

الجمهورية الجزائرية الديمقراطية الشعبية
République algérienne démocratique et populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique
جامعة عين تموشنت بلحاج بوشعيب
Université –Ain Temouchent- Belhadj Bouchaib
Faculté des Sciences et de Technologie
Département Génie Mécanique



Projet de Fin d'Etudes
Pour l'obtention du diplôme de Master en : Génie Mécanique
Domaine : Sciences et Technologies
Filière : Génie Mécanique
Spécialité : Énergétique
Thème

Prévision de la demande électrique journalière d'une maison à l'aide de modèles de séries temporelles et de Machine Learning

Présenté Par :

- 1)REBHI Rim Chames El Houda
- 2)BELABANE Hayet Kaouter Zahra

Devant le jury composé de :

Pr. NEHARI Taieb	UAT.B.B .(Ain Temouchent)	Président
Dr. REMLAOUI Ahmed	UAT.B.B .(Ain Temouchent)	Examineur
Dr. DORBANE Abdelhakim	UAT.B.B.(Ain Temouchent)	Encadrant

Année Universitaire 2025/2026

Remerciements

Avant tout, nous remercions Dieu Tout-Puissant de nous avoir donné la santé, la force et la persévérance nécessaires pour mener à bien ce travail.

*Nous tenons à exprimer notre profonde gratitude et nos sincères remerciements à notre encadreur, **Monsieur Dorbane Abdelhakim**, pour son accompagnement, sa disponibilité, ses précieux conseils et ses orientations pertinentes tout au long de la réalisation de ce projet de fin d'études. Ses compétences scientifiques, son soutien et sa confiance ont été d'une grande valeur pour l'aboutissement de ce travail.*

Nous adressons également nos vifs remerciements à l'ensemble des enseignants du département, qui ont contribué à notre formation et nous ont transmis leur savoir et leur expérience tout au long de notre parcours universitaire.

Enfin, nous exprimons notre reconnaissance à toutes les personnes qui, de près ou de loin, ont participé à la réalisation de ce mémoire et ont contribué à la réussite de ce travail

Dédicace

Tout d'abord, je tiens à remercier DIEU de m'avoir donné la force et le courage de mener à bien ce modeste travail.

À ma mère, Soumia.

Merci pour ton amour inconditionnel, ta patience infinie et ton soutien . Tu as été ma lumière dans l'ombre, ma force dans les épreuves, et chaque réussite que j'ai connue porte ton empreinte.

À mon père, Abd EL Hadi.

Merci pour ton courage et la sagesse que tu m'as transmise. Tu m'as appris que la vraie valeur se construit dans l'effort, la constance et la capacité de se relever après chaque difficulté.

À mes frères, Ibrahim, Moustapha ,Badr Ed doudja , Mohamed et Yacine.

Merci pour votre affection, votre joie de vivre et tous les beaux moments partagés ensemble. Votre présence dans ma vie est une source de bonheur et de motivation. Je vous souhaite un avenir rempli de réussite, de bonheur et d'épanouissement.

À mes enseignants et mon encadrant Dr. DORBANE Abdelhakim.

Je tiens à vous exprimer ma profonde gratitude pour votre accompagnement, votre disponibilité et la qualité de votre encadrement. Vos conseils précieux, votre rigueur et votre soutien ont largement contribué à la réalisation de ce travail. Merci pour le savoir que vous m'avez transmis et pour l'inspiration que vous m'avez apportée tout au long de mon parcours.

À mes amis proches, Fatma Et Houda.

Merci pour vos mots, votre écoute et votre réconfort dans les moments d'incertitude et de fatigue. Votre présence m'a aidée à garder le cap et à rester fidèle à mes objectifs. À toutes les personnes qui ont contribué, de près ou de loin, à mon parcours et à la réalisation de ce mémoire, je dédie ce travail avec toute ma reconnaissance et mon affection.

REBHI Rim Chames El Houda

Dédicace

C'est avec une immense joie et une profonde émotion que je dédie ce modeste travail à ceux qui occupent une place précieuse dans mon cœur.

À mon cher père, Kamel,

toi qui as toujours été un exemple de courage, de générosité et de sagesse. Merci pour tous les sacrifices consentis, pour ton soutien inconditionnel et pour la confiance que tu m'as toujours accordée. Que ce travail soit le témoignage de mon amour, de mon respect et de mon éternelle reconnaissance.

À ma très chère mère, Kamla,

source de tendresse, d'affection et de bienveillance. Tes prières, tes encouragements et ton amour ont toujours été ma plus grande force. Aucun mot ne saurait exprimer toute la gratitude et l'amour que je te porte. Puisse Dieu te préserver et te combler de bonheur.

À mon cher frère, Mohamed, et à ma chère Jinen,

pour leur affection, leur présence et leur soutien tout au long de mon parcours. Je leur souhaite une vie remplie de bonheur, de réussite et de sérénité.

À mes chères tantes,

pour leur amour, leur attention et leurs encouragements qui m'ont toujours accompagnée. Qu'elles trouvent dans ce travail l'expression de ma profonde affection et de ma reconnaissance.

À mes très chères amies, Souhila et Ferial,

plus que des amies, des sœurs de cœur. Merci pour votre présence, votre soutien, vos encouragements et tous les moments de joie et de partage qui ont rendu ce parcours plus beau et plus agréable. Votre amitié sincère a été une source de réconfort et de motivation. Je vous souhaite tout le bonheur, la réussite et l'épanouissement que vous méritez.

Enfin, je dédie ce travail à toutes les personnes qui m'ont soutenue, encouragée et accompagnée tout au long de mon parcours.

Belabane hayet zahra kaoutar

Résumé

Ce projet de fin d'études de Master porte sur la prévision de la consommation résidentielle d'électricité à l'aide des techniques d'apprentissage automatique (Machine Learning). Son principal objectif est de développer des modèles prédictifs capables d'estimer avec précision la demande en électricité à partir des données historiques de consommation et de variables explicatives pertinentes. Après une revue approfondie de la littérature, quatre modèles de prédiction — la régression linéaire, Random Forest, XGBoost et CatBoost — ont été développés, implémentés et comparés. Les performances de ces modèles ont été évaluées à l'aide des indicateurs suivants : l'erreur absolue moyenne (MAE), l'erreur quadratique moyenne (MSE), la racine de l'erreur quadratique moyenne (RMSE), l'erreur absolue moyenne en pourcentage (MAPE) et le coefficient de détermination (R^2). Les résultats expérimentaux démontrent l'efficacité des techniques d'apprentissage automatique pour améliorer la précision des prévisions. Le meilleur modèle a obtenu une MSE de 0,408103, une RMSE de 0,638829, une MAE de 0,480083 et un coefficient de détermination R^2 de 0,986685, témoignant d'excellentes performances prédictives. Ces résultats mettent en évidence le potentiel du Machine Learning pour favoriser une gestion efficace de l'énergie, améliorer la prévision de la demande en électricité et optimiser le fonctionnement des réseaux électriques.

Abstract

This Master's end-of-study project focuses on forecasting residential electricity consumption using Machine Learning techniques. The primary objective is to develop predictive models capable of accurately estimating electricity demand based on historical consumption data and relevant explanatory variables. Following a comprehensive literature review, four prediction models—Linear Regression, Random Forest, XGBoost, and CatBoost—were developed, implemented, and comparatively evaluated. The performance of these models was assessed using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). The experimental results demonstrate the effectiveness of machine learning techniques in improving forecasting accuracy. The best-performing model achieved an MSE of **0.408103**, an RMSE of **0.638829**, an MAE of **0.480083**, and an R^2 score of 0.986685, indicating excellent predictive performance. These findings highlight the potential of machine learning to support efficient energy management, accurate electricity demand forecasting, and improved electrical grid operation.

ملخص

تركز هذه المذكرة على التنبؤ باستهلاك الكهرباء في القطاع السكني باستخدام تقنيات التعلم الآلي. ويتمثل الهدف الرئيسي في تطوير نماذج تنبؤية قادرة على تقدير الطلب على الطاقة الكهربائية بدقة اعتماداً على بيانات الاستهلاك التاريخية والمتغيرات التفسيرية ذات الصلة. وبعد إجراء مراجعة شاملة للأدبيات العلمية، تم تطوير وتنفيذ ومقارنة أربعة نماذج للتنبؤ، وهي: الانحدار الخطي، والغابة العشوائية (Random Forest)، وXGBoost، وCatBoost. وقد تم تقييم أداء هذه النماذج باستخدام مؤشرات متوسط الخطأ المطلق (MAE)، ومتوسط مربع الخطأ (MSE)، والجذر التربيعي لمتوسط مربع الخطأ (RMSE)، ومتوسط نسبة الخطأ المطلق (MAPE)، ومعامل التحديد (R^2). وأظهرت النتائج التجريبية أن تقنيات التعلم الآلي تحقق دقة عالية في التنبؤ، حيث حقق أفضل نموذج قيمة $R^2 = 0.986685$ ، و $MSE = 0.408103$ ، و $RMSE = 0.638829$ ، و $MAE = 0.480083$ ، ومعامل تحديد $R^2 = 0.986685$ ، مما يعكس قدرة تنبؤية ممتازة. وتبرز هذه النتائج فعالية تقنيات التعلم الآلي في تحسين دقة التنبؤ باستهلاك الكهرباء، بما يساهم في دعم الإدارة الرشيدة للطاقة، وتحسين التنبؤ بالطلب، وتعزيز كفاءة تشغيل الشبكات الكهربائية.

Liste des figures

Figure 1 : Fonctionnement d'une centrale thermique et d'une centrale hydroélectrique [9]	6
Figure 2 : Evolution de la facture d'électricité en fonction des températures en ile-de-France (Mai 2020 – Avril 2021, comparée a Mai 2019 – Avril 2021) [13]	8
Figure 3 : la consommation d'Energie, en hiver, après correction des effets météorologiques[20]	9
Figure 4: comparaison des moyennes mobiles pour l'extraction de la tendance [31]	12
Figure 5 : composante saisonnière de la série temporelle (1949 -1960) [31]	13
Figure 6:Analyse de la décomposition saisonnière d'une série temporelle [34] [31]	13
Figure 7: Le fonctionnement d'un algorithme de machine learning [38]	14
Figure 8: Architecture classique d'un réseau de neurones artificiels [43].	18
Figure 9 : Consommation d'électricité par secteur en Algérie 2021 (APRUE 2022)[46]	19
Figure 10:Schéma du modèle hybride SARIMAX-LSTM pour la prévision de la demande énergétique d'un campus universitaire [61]	24
Figure 11: Architecture du Framework de prévision de la consommation électrique [66].....	26
Figure 12: Régression linéaire [70].....	30
Figure 13: Random forest [72].....	31
Figure 14 : XGBoost Algorithm – Objective [74]	32
Figure 15 :The flow diagram of the CatBoost model[75].....	33
Figure 16 : Image réelle du programme de la simulation de notre étude.....	39
Figure 17 : Violin plot de Has_AC	41
Figure 18 : Violin plot de Avg_Temperature_C	42
Figure 19 : Violin plot de Household_Size.....	43
Figure 20 : Violin plots des variables numériques.....	44
Figure 21: Matrice de corrélation des sept paramètres originaux	45
Figure 22: Organigramme du cadre de prédiction de notre étude.....	46
Figure 23 : Comparaison des coefficients de détermination des modèles utilisés	48
Figure 24 : Comparaison des erreurs (MSE, RMSE et MAE) sous forme d'histogramme groupé.....	49
Figure 25 : Gain de performance des modèles de Machine Learning par rapport à la régression linéaire	50
Figure 26 : Résultats du SHAP du modèle XGBoost	52

Listes des tableaux

Tableau 1 - Household Energy Consumption - April 2025 (90,000 Records).....	29
Tableau 2 - Évaluation comparative des performances de quatre modèles	47

Sommaire

Résumé.....	1
Abstract	1
ملخص.....	1
Liste des figures	2
Listes des tableaux	3
Sommaire	4
Introduction Générale	1
Chapitre I: Généralités	4
1.1 Généralités sur l'énergie électrique	5
1.1.1 Définition de l'énergie électrique	5
1.1.2 Production et distribution de l'électricité.....	5
1.2 Consommation énergétique dans les habitations	7
1.3 Facteurs influençant la demande électrique	7
1.3.1 Température et conditions climatiques	7
1.3.2 Comportement des occupants	8
1.3.3 Saisons et périodes de la journée	8
1.4 Importance de la prévision de la demande énergétique	9
1.4.1 Optimisation de la consommation	10
1.4.2 Gestion des réseaux électriques	10
1.4.3 Contribution à l'efficacité énergétique	10
1.4.4 Notions de base sur les séries temporelles	11
1.5 Introduction au Machine Learning.....	14
1.5.1 Définition du Machine Learning.....	14
1.6 Modèles de Machine Learning utilisés pour la prévision	15
1.7 Conclusion	15
Chapitre II : Recherche bibliographique.....	16
2.1 Introduction.....	17
2.2 Les études précédentes.....	17
2.3 Conclusion	26
Chapitre III : Méthodologie de la recherche	27
3.1 Introduction.....	28
3.2 Description des paramètres d'entrée et de sortie	28
3.3 Présentation des différents modèles de prédiction utilisant l'IA	30
3.3.1 Régression linéaire:.....	30
3.3.2 Modèle XGBoost	32

3.3.3 Modèle CatBoost	33
3.4 Prétraitement des données.....	34
3.5 Nettoyage des données.....	34
3.6 Transformation des variables temporelles	34
3.7 Normalisation et mise à l'échelle.....	34
3.8 Séparation des données (Train/Test).....	35
3.9 Sélection des variables.....	35
3.10 Évaluation des performances des modèles	35
3.10.1 Erreur absolue moyenne (MAE).....	35
3.10.2 Erreur quadratique moyenne (RMSE)	35
3.10.3 Coefficient de détermination (R^2).....	36
3.10.4 Analyse comparative des modèles	36
3.11 Conclusion	36
Chapitre IV: Résultats et discussions.....	37
4.1 Introduction.....	38
4.2 Présentation de Python.....	38
4.3 Résultats de la prévision	40
4.4 Résultats du tracage du diagramme SHAP	50
4.5 Conclusion	52
Conclusion Générale	53
Références.....	56

Introduction Générale

L'énergie électrique constitue l'un des éléments fondamentaux du développement économique, industriel et social. Elle est devenue indispensable dans la vie quotidienne en raison de son utilisation croissante dans les secteurs résidentiel, industriel et tertiaire. Avec l'augmentation de la population, l'urbanisation accélérée et la multiplication des équipements électriques, la demande en énergie connaît une croissance continue, ce qui engendre de nouveaux défis liés à la production, au transport, à la distribution et à la gestion efficace des ressources énergétiques.

Le secteur résidentiel représente une part importante de la consommation électrique mondiale. Cette consommation est influencée par plusieurs facteurs, notamment les conditions climatiques, les caractéristiques des habitations, le nombre d'occupants ainsi que leurs habitudes de consommation. Les variations de ces paramètres rendent le comportement de la demande énergétique complexe et difficile à modéliser. Par conséquent, la prévision de la consommation électrique est devenue une problématique majeure pour les fournisseurs d'énergie et les gestionnaires des réseaux électriques, car elle permet d'assurer un équilibre entre la production et la consommation, de réduire les coûts d'exploitation et d'améliorer l'efficacité énergétique.

Traditionnellement, plusieurs méthodes statistiques ont été utilisées pour prévoir la demande énergétique. Toutefois, ces méthodes présentent certaines limites lorsqu'il s'agit de modéliser des phénomènes complexes et non linéaires. Grâce aux avancées récentes de l'intelligence artificielle, et plus particulièrement du Machine Learning, de nouvelles approches ont été développées afin d'exploiter efficacement les données historiques et d'obtenir des prévisions plus précises. Ces techniques permettent de mettre en évidence les relations existantes entre les différentes variables influençant la consommation énergétique et d'améliorer considérablement les performances des modèles de prédiction.

Dans ce contexte, ce travail de recherche s'intéresse à la prévision de la consommation électrique résidentielle à partir d'un ensemble de données décrivant les caractéristiques des ménages et les paramètres environnementaux. L'objectif principal est de développer et de comparer plusieurs modèles de Machine Learning, à savoir la Régression Linéaire, Random Forest, XGBoost et CatBoost, afin d'identifier le modèle le plus performant pour la prédiction de la consommation énergétique. L'évaluation des modèles est réalisée à l'aide de différentes métriques telles que l'erreur absolue moyenne (MAE), l'erreur quadratique

moyenne (MSE), la racine de l'erreur quadratique moyenne (RMSE) et le coefficient de détermination (R^2).

Au-delà de l'aspect prédictif, cette étude vise également à mettre en évidence l'importance des méthodes d'intelligence artificielle dans le domaine de la gestion énergétique. Une prévision fiable de la demande électrique permet non seulement d'optimiser la consommation d'énergie et de réduire les pertes, mais également de favoriser le développement des habitations intelligentes et des réseaux électriques modernes (Smart Grids), contribuant ainsi à une utilisation plus rationnelle et durable des ressources énergétiques.

Ce mémoire est organisé en quatre chapitres. Le premier chapitre présente les généralités sur l'énergie électrique, la consommation énergétique résidentielle, les séries temporelles ainsi que les notions fondamentales du Machine Learning. Le deuxième chapitre est consacré à une étude bibliographique des travaux antérieurs portant sur la prévision de la consommation énergétique. Le troisième chapitre décrit la méthodologie adoptée, les données utilisées, les étapes de prétraitement et les différents modèles de prédiction développés. Enfin, le quatrième chapitre présente les résultats obtenus, leur interprétation ainsi qu'une comparaison des performances des modèles étudiés, avant de conclure sur les principales contributions de ce travail et les perspectives de recherche futures.

Chapitre I: Généralités

1.1 Généralités sur l'énergie électrique

1.1.1 Définition de l'énergie électrique

L'énergie électrique est une forme d'énergie résultant du déplacement des charges électriques dans un conducteur sous l'effet d'une différence de potentiel. Elle constitue aujourd'hui l'une des formes d'énergie les plus utilisées dans le monde en raison de sa facilité de transport, de transformation et d'utilisation dans de nombreux domaines tels que l'industrie, les transports, les services et le secteur résidentiel [1].

L'énergie électrique est généralement mesurée en kilowattheure (kWh), qui représente la quantité d'énergie consommée par un appareil de puissance d'un kilowatt fonctionnant pendant une heure. Cette unité est couramment utilisée pour la facturation de l'électricité dans les habitations. Grâce à sa flexibilité et à son efficacité, l'électricité joue un rôle essentiel dans le développement économique et social ainsi que dans l'amélioration du confort de vie des populations.

1.1.2 Production et distribution de l'électricité

La production d'électricité consiste à transformer différentes formes d'énergie primaire en énergie électrique à l'aide de centrales de production. Selon la source d'énergie utilisée, plusieurs types de centrales peuvent être distingués, notamment les centrales thermiques, hydrauliques, nucléaires et les centrales basées sur les énergies renouvelables comme le solaire photovoltaïque et l'éolien [2], [3] [4], [5].

Dans les centrales thermiques, l'électricité est produite en brûlant des combustibles fossiles tels que le charbon, le gaz naturel ou le pétrole afin de générer de la vapeur qui entraîne une turbine couplée à un alternateur [6], [7], [8]. Les centrales hydrauliques utilisent quant à elles la force de l'eau stockée dans des barrages pour faire tourner des turbines [9], le schéma du fonctionnement d'une centrale thermique et d'une centrale hydroélectrique est montré dans la Figure 1 [10]. Par ailleurs, les technologies renouvelables, comme les panneaux photovoltaïques, permettent de convertir directement l'énergie solaire en énergie électrique.

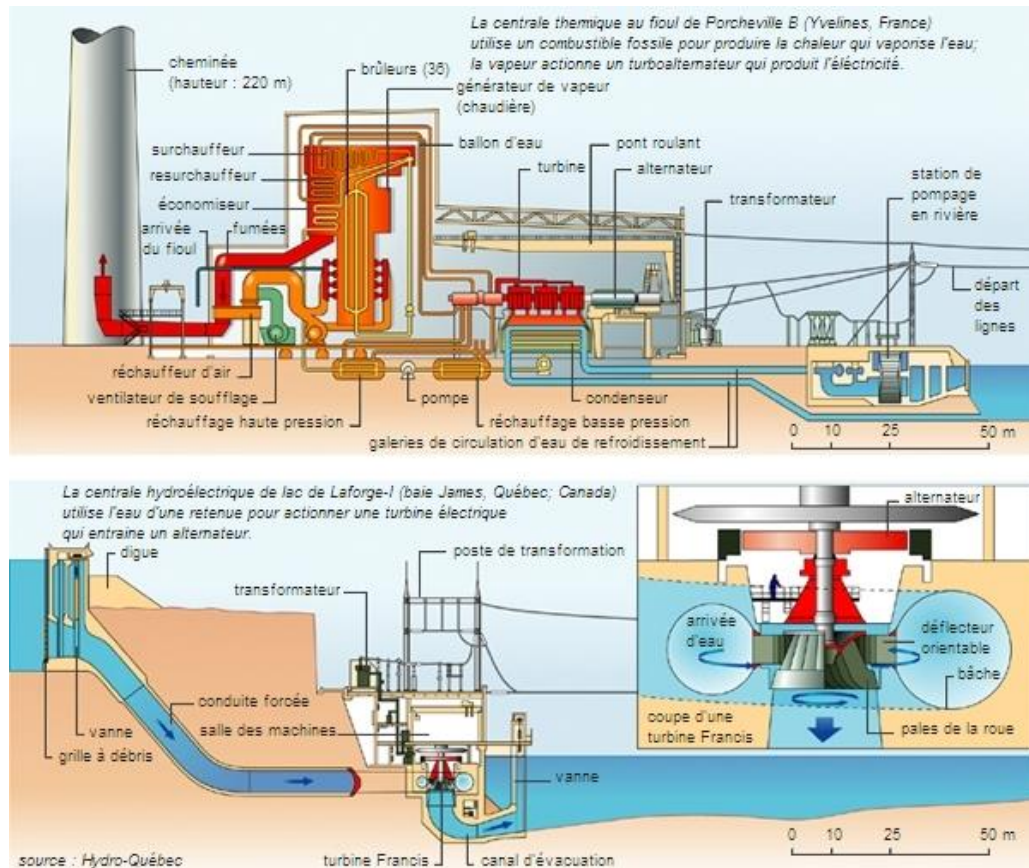


Figure 1 : Fonctionnement d'une centrale thermique et d'une centrale hydroélectrique [11]

Une fois produite, l'électricité est acheminée vers les consommateurs à travers un réseau complexe appelé réseau électrique. Ce réseau comprend trois étapes principales : le transport, la distribution et la consommation. Tout d'abord, l'électricité produite dans les centrales est transportée sur de longues distances à l'aide de lignes à haute tension, ce qui permet de réduire les pertes d'énergie lors du transport. Ensuite, elle passe par des postes de transformation où la tension est réduite afin d'être adaptée aux réseaux de distribution locaux. Enfin, l'électricité est distribuée aux utilisateurs finaux, notamment les habitations, les entreprises et les infrastructures publiques, via des réseaux de moyenne et basse tension. Dans le secteur résidentiel, l'électricité alimente une grande variété d'appareils et de systèmes, ce qui rend la gestion et la prévision de la consommation énergétique particulièrement importantes pour améliorer l'efficacité énergétique et optimiser l'utilisation des ressources.

1.2 Consommation énergétique dans les habitations

La consommation d'électricité dans les habitations représente une part importante de la demande énergétique globale. Elle dépend de plusieurs facteurs tels que le nombre d'occupants, les conditions climatiques, la taille du logement et les habitudes d'utilisation des équipements électriques.

Dans une maison, certains appareils consomment une grande partie de l'électricité. Parmi les plus importants figurent les systèmes de chauffage et de climatisation, les appareils électroménagers (réfrigérateur, lave-linge, four, etc.), l'éclairage, ainsi que les appareils électroniques comme les téléviseurs et les ordinateurs [12]. L'utilisation croissante des équipements domestiques contribue à l'augmentation de la consommation énergétique.

Le profil de consommation d'une maison représente l'évolution de la demande électrique au cours du temps. Généralement, la consommation est faible pendant la nuit, augmente le matin avec les activités quotidiennes, diminue légèrement durant la journée, puis atteint un pic le soir lorsque les occupants utilisent plusieurs appareils simultanément. L'analyse de ce profil permet de mieux comprendre les habitudes de consommation et d'améliorer la gestion de l'énergie dans les habitations [13].

1.3 Facteurs influençant la demande électrique

La demande électrique dans les habitations n'est pas constante et varie en fonction de plusieurs facteurs liés à l'environnement, au comportement des occupants et aux conditions temporelles. L'identification de ces facteurs est essentielle pour comprendre les variations de la consommation énergétique et pour développer des modèles fiables de prévision de la demande électrique.

1.3.1 Température et conditions climatiques

La température extérieure et les conditions climatiques jouent un rôle déterminant dans la consommation d'électricité d'une habitation. Lorsque la température est très basse, les systèmes de chauffage sont davantage sollicités afin de maintenir une température intérieure confortable. À l'inverse, lors des périodes de forte chaleur, l'utilisation des systèmes de climatisation et de ventilation augmente considérablement la demande électrique[14]. La Figure 2 présente l'évolution de la facture d'électricité en fonction des températures en Ile-de-France (Mai 2020 – Avril 2021, comparée à Mai 2019 – Avril 2021), montrant les pics

d'utilisation de l'électricité pendant l'été, ceci peut être dû à cause de l'utilisation immense du climatiseur et d'autres appareils électriques.

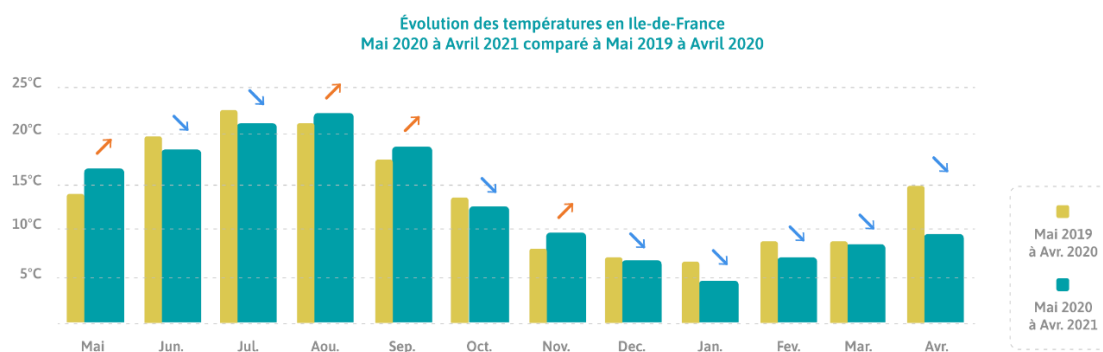


Figure 2 : Evolution de la facture d'électricité en fonction des températures en ile-de-France (Mai 2020 – Avril 2021, comparée à Mai 2019 – Avril 2021) [15]

D'autres conditions climatiques, telles que l'humidité, l'ensoleillement ou encore la vitesse du vent, peuvent également influencer la consommation énergétique [16]. Par exemple, un fort ensoleillement peut réduire les besoins en éclairage durant la journée, tandis que des conditions météorologiques extrêmes peuvent entraîner une augmentation significative de la consommation électrique.

1.3.2 Comportement des occupants

Le comportement des occupants constitue un facteur important dans la variation de la consommation électrique domestique [17]. Les habitudes de vie, les horaires de présence dans le logement et les modes d'utilisation des appareils électriques influencent directement la demande énergétique.

Par exemple, l'utilisation fréquente d'appareils électroménagers, la durée d'utilisation de l'éclairage ou encore le réglage de la température des systèmes de chauffage et de climatisation peuvent entraîner des différences importantes dans la consommation d'électricité. De plus, le nombre de personnes vivant dans une maison a également un impact direct sur la demande énergétique.

1.3.3 Saisons et périodes de la journée

La demande électrique varie également en fonction des saisons et des moments de la journée. En hiver, la consommation d'électricité augmente généralement en raison de l'utilisation

accrue des systèmes de chauffage et de l'éclairage. En été, la demande peut également augmenter à cause de l'utilisation des climatiseurs et des ventilateurs [18].

À l'échelle quotidienne, la consommation électrique suit généralement un profil caractéristique [19], [20]. Elle est faible pendant la nuit, augmente le matin avec le début des activités domestiques, peut diminuer durant la journée lorsque les occupants sont absents, puis atteint souvent un pic en soirée lorsque plusieurs appareils sont utilisés simultanément.

La Figure 3 présente la consommation d'Énergie, en hiver, montrant la variation de la consommation en fonction des mois de l'année [21].

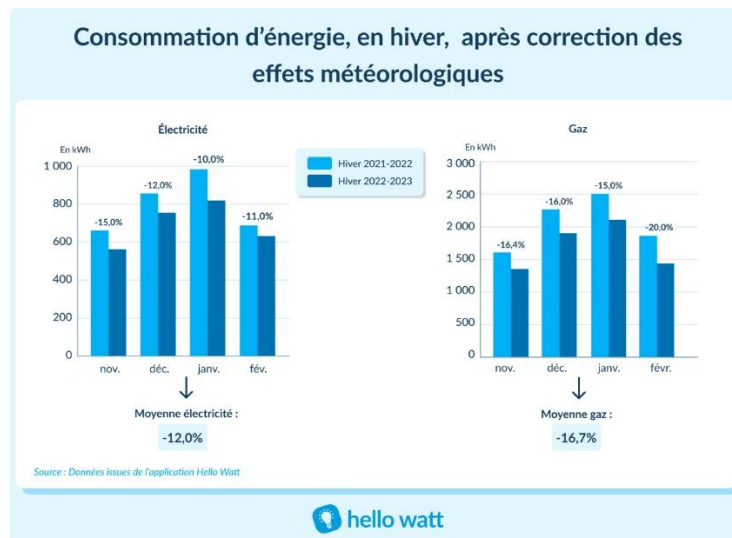


Figure 3 : la consommation d'Énergie, en hiver, après correction des effets météorologiques[22]

1.4 Importance de la prévision de la demande énergétique

La prévision de la demande énergétique constitue un élément essentiel dans la gestion moderne des systèmes électriques. Elle permet d'anticiper les besoins en électricité sur différentes périodes de temps, allant de quelques heures à plusieurs jours ou mois. Une estimation précise de la consommation future facilite la planification de la production et de la distribution de l'électricité, tout en contribuant à une utilisation plus efficace des ressources énergétiques. Dans le contexte des habitations intelligentes et des réseaux électriques modernes, la prévision de la demande énergétique joue un rôle de plus en plus important pour améliorer la performance des systèmes énergétiques et réduire les pertes.

1.4.1 Optimisation de la consommation

La prévision de la demande énergétique permet d'optimiser l'utilisation de l'électricité dans les habitations [23]. En anticipant les périodes de forte consommation, il devient possible d'adapter l'utilisation des appareils électriques afin de réduire les pics de demande et d'améliorer la gestion de l'énergie.

Par exemple, certains appareils électroménagers peuvent être programmés pour fonctionner pendant les périodes où la demande est plus faible. Cette approche contribue à réduire les coûts énergétiques pour les consommateurs et à limiter la surcharge des systèmes électriques. De plus, dans le cadre des maisons intelligentes, les systèmes de gestion énergétique peuvent utiliser les prévisions de consommation pour automatiser le fonctionnement de certains équipements.

1.4.2 Gestion des réseaux électriques

La prévision de la demande énergétique est également essentielle pour la gestion des réseaux électriques[24]. Les opérateurs des réseaux doivent équilibrer en permanence la production et la consommation d'électricité afin d'assurer la stabilité du système. Une estimation précise de la demande permet d'optimiser la planification de la production d'électricité et d'éviter les situations de surcharge ou de pénurie énergétique [25]. Elle facilite également l'intégration des sources d'énergie renouvelables, telles que l'énergie solaire et éolienne, dont la production peut être variable et dépendante des conditions climatiques.

1.4.3 Contribution à l'efficacité énergétique

Enfin, la prévision de la demande énergétique contribue de manière significative à l'amélioration de l'efficacité énergétique[26]. En identifiant les périodes de forte consommation et les habitudes d'utilisation de l'électricité, il devient possible de mettre en place des stratégies visant à réduire le gaspillage énergétique. Ces stratégies peuvent inclure l'optimisation du fonctionnement des appareils électriques, l'amélioration de l'isolation des bâtiments ou encore l'utilisation de technologies intelligentes pour gérer la consommation énergétique. Ainsi, la prévision de la demande énergétique constitue un outil important pour promouvoir une utilisation plus rationnelle et durable de l'électricité dans le secteur résidentiel.

1.4.4 Notions de base sur les séries temporelles

a) Définition d'une série temporelle :

Une série temporelle est une suite d'observations numériques ordonnées dans le temps, généralement relevées à intervalles réguliers (heure, jour, mois, etc.[27], [28]).

Dans le contexte énergétique, une série temporelle peut représenter l'évolution de la consommation électrique d'une habitation au fil du temps.

L'analyse des séries temporelles vise à comprendre la structure des données passées afin de modéliser leur comportement et de prévoir leurs valeurs futures[29], [30]. Cette approche est largement utilisée dans le domaine de l'énergie pour anticiper la demande électrique et optimiser la gestion des ressources.

b) Composantes d'une série temporelle :

Une série temporelle est généralement constituée de plusieurs composantes fondamentales qui permettent de mieux comprendre sa structure et son comportement dans le temps.[31]

L'identification et l'analyse de ces différentes composantes sont essentielles pour appliquer des méthodes de modélisation et de prévision adaptées.

c) La tendance (Trend) :

La tendance représente l'évolution générale de la série sur une longue période. Elle traduit la direction globale des données, qu'elle soit croissante, décroissante ou stable. Cette composante est souvent liée à des facteurs structurels tels que la croissance démographique, le développement économique ou l'évolution technologique. Par exemple, dans le cas de la consommation électrique, on observe généralement une tendance à la hausse due à l'augmentation du nombre d'appareils électriques et à l'amélioration du niveau de vie.[32]

L'identification de la tendance permet de mieux comprendre le comportement global de la série et d'anticiper son évolution future.

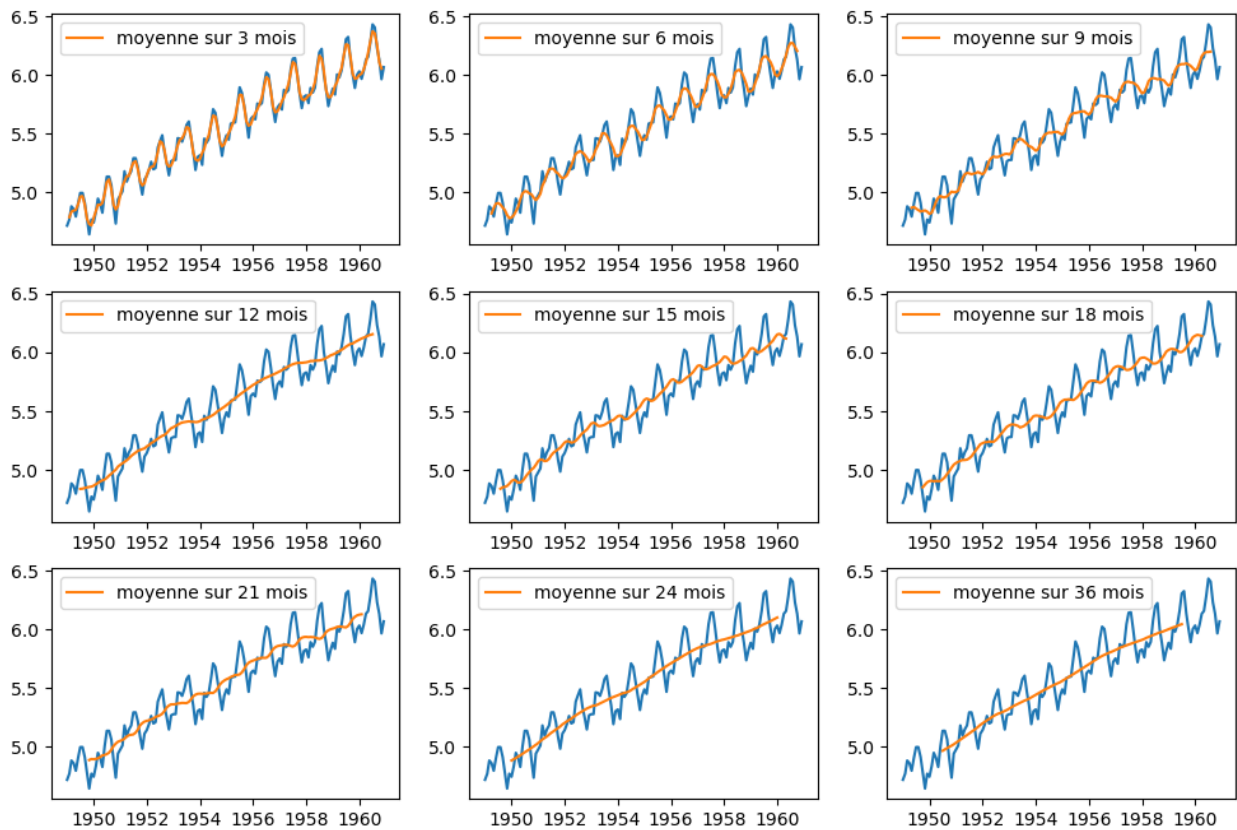


Figure 4: comparaison des moyennes mobiles pour l'extraction de la tendance [33]

La saisonnalité (Seasonality) : La saisonnalité correspond aux variations régulières et périodiques qui se répètent à intervalles fixes (journaliers, hebdomadaires, mensuels ou annuels). Ces fluctuations sont souvent influencées par des facteurs climatiques, sociaux ou comportementaux. Par exemple, la consommation d'électricité peut augmenter en été en raison de l'utilisation intensive des climatiseurs, ou en hiver à cause du chauffage[34]. La prise en compte de la saisonnalité est essentielle pour améliorer la précision des modèles de prévision, notamment dans les séries liées à l'énergie.

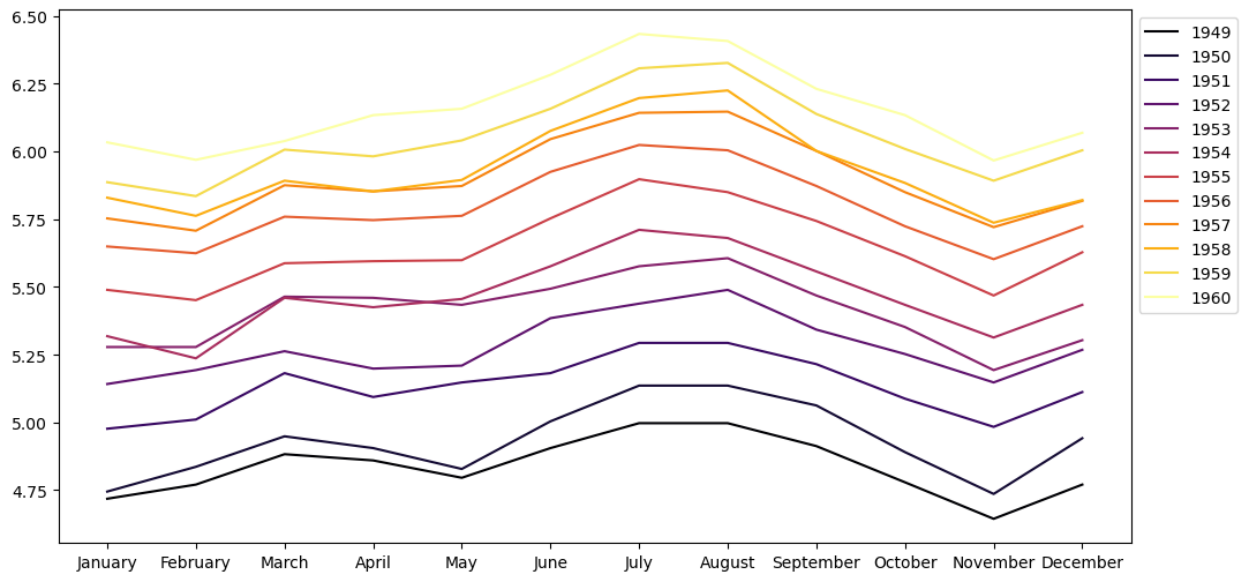


Figure 5 : composante saisonnière de la série temporelle (1949 -1960) [33]

Le bruit (Noise) : Le bruit représente la composante aléatoire de la série temporelle. Il s'agit de fluctuations imprévisibles qui ne suivent aucune structure identifiable. Ces variations peuvent être dues à des événements exceptionnels, des erreurs de mesure ou des facteurs externes non contrôlés. Bien que le bruit ne puisse pas être modélisé directement, il est important de le minimiser ou de le filtrer afin d'améliorer la qualité des prévisions [35].

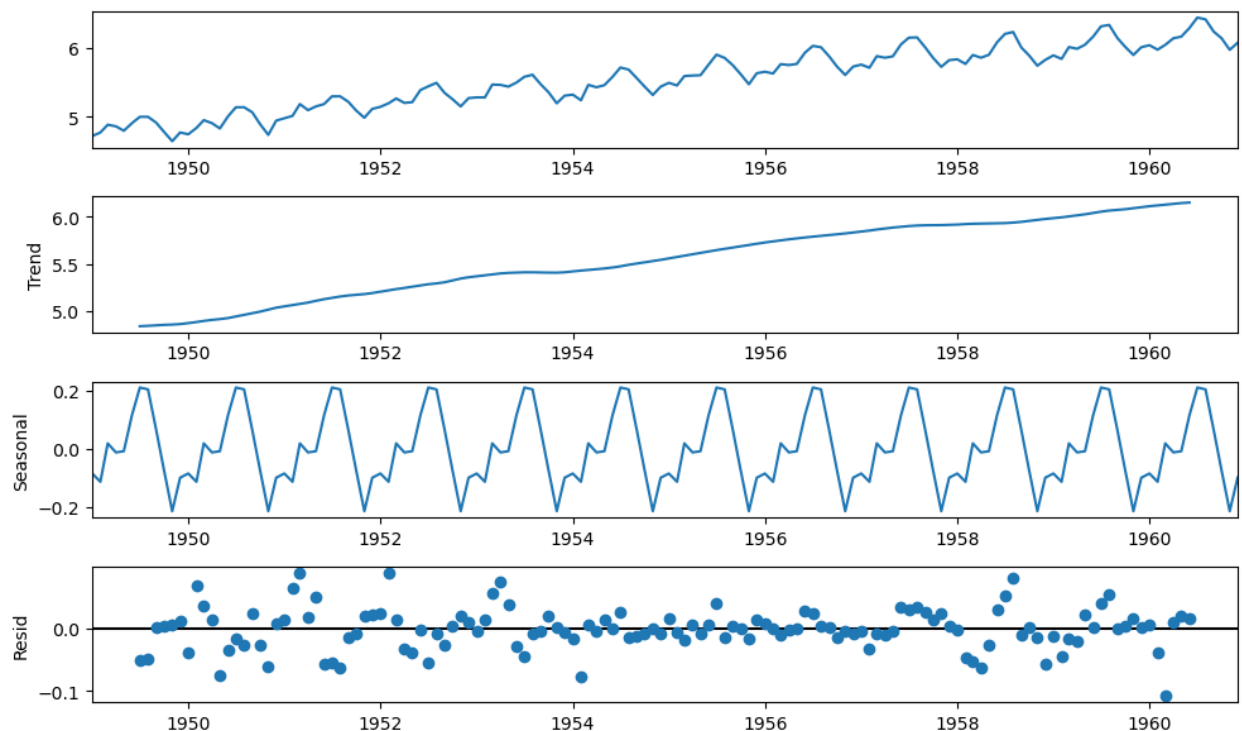


Figure 6: Analyse de la décomposition saisonnière d'une série temporelle [36] [33]

La composante cyclique (Cycle) : La composante cyclique correspond à des variations à long terme qui ne sont pas strictement périodiques, contrairement à la saisonnalité [37]. Ces cycles peuvent être influencés par des facteurs économiques, politiques ou sociaux, tels que les crises économiques ou les changements de politiques énergétiques. Leur durée est généralement plus longue et irrégulière, ce qui les rend plus difficiles à identifier et à modéliser [38].

1.5 Introduction au Machine Learning

1.5.1 Définition du Machine Learning

Le Machine Learning (apprentissage automatique) est une branche de l'intelligence artificielle qui permet aux machines de apprendre à partir des données et d'améliorer leurs performances au fil du temps, sans être explicitement programmées pour effectuer une tâche spécifique[39]. Plutôt que de suivre des règles fixes, les algorithmes de Machine Learning identifient des patterns et des relations complexes dans les données pour générer des prédictions ou prendre des décisions.

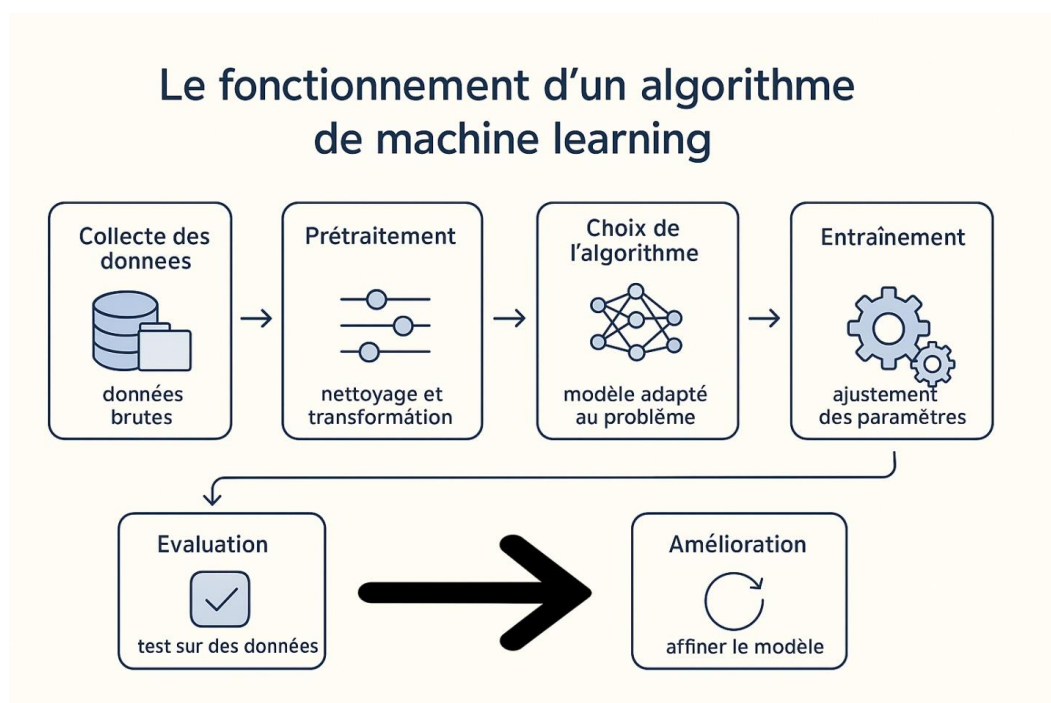


Figure 7: Le fonctionnement d'un algorithme de machine learning [40]

Dans le contexte de la prévision de la demande électrique, le Machine Learning offre des capacités avancées pour modéliser des comportements énergétiques souvent complexes et dynamiques. Il permet de construire des modèles capables de détecter des relations non linéaires entre différentes variables explicatives, telles que :

- La consommation électrique historique,
- Les conditions météorologiques (température, humidité, ensoleillement...),
- Le temps ou le calendrier (heures de la journée, jours de la semaine, saisons),
- Le comportement des utilisateurs et les habitudes de consommation.

1.6 Modèles de Machine Learning utilisés pour la prévision

Dans le domaine de la prévision de la demande électrique, les modèles de Machine Learning sont particulièrement utiles pour exploiter les relations complexes et non linéaires entre la consommation d'énergie et les différentes variables explicatives[41]. Ces modèles permettent de mieux anticiper les fluctuations de la demande, d'optimiser la gestion énergétique et de soutenir des stratégies de réduction des coûts et d'efficacité énergétique. [42]

1.7 Conclusion

Dans ce chapitre, nous avons présenté les notions fondamentales relatives à la consommation énergétique et à son importance dans le contexte actuel de gestion des ressources électriques. Nous avons également abordé les principaux facteurs influençant la demande énergétique ainsi que les concepts liés aux séries temporelles et aux méthodes de prévision. Par ailleurs, les bases du Machine Learning et les différentes approches utilisées pour la modélisation et la prédiction ont été introduites. Ces éléments théoriques ont permis de mettre en évidence l'intérêt des techniques d'intelligence artificielle dans l'amélioration de la précision des prévisions de consommation électrique. Ainsi, ce cadre conceptuel constitue une base essentielle pour la suite de ce travail. Le chapitre suivant sera consacré à l'étude des travaux de recherche antérieurs portant sur la prévision de la consommation énergétique, afin de mieux situer notre contribution par rapport aux approches déjà proposées dans la littérature.

Chapitre II : Recherche bibliographique

Dans Ce chapitre présente les travaux antérieurs portant sur la prévision de la consommation énergétique ainsi que les notions fondamentales relatives aux séries temporelles et au machine learning. Il permet de situer le présent travail par rapport aux méthodes existantes

2.1 Introduction

La prévision de la consommation électrique constitue un domaine de recherche particulièrement actif en raison de son importance dans la gestion des réseaux énergétiques, l'optimisation de la consommation et l'intégration des sources d'énergie renouvelables. Au cours des dernières années, de nombreuses études ont été menées afin de développer des modèles capables d'estimer avec précision la demande énergétique à partir de données historiques, météorologiques et comportementales. L'évolution des technologies de l'information et l'augmentation de la disponibilité des données ont favorisé l'émergence de nouvelles approches de prévision basées sur l'apprentissage automatique (Machine Learning) et l'apprentissage profond (Deep Learning). Ces méthodes offrent une capacité supérieure à modéliser les relations complexes et non linéaires entre les variables influençant la consommation électrique, dépassant souvent les performances des approches statistiques traditionnelles. Dans ce contexte, plusieurs chercheurs ont proposé et comparé différentes techniques de prévision, notamment les modèles statistiques, les réseaux de neurones artificiels, les machines à vecteurs de support, les méthodes de boosting, les modèles hybrides ainsi que les architectures de Deep Learning. Les résultats obtenus mettent en évidence l'importance du choix des données d'entrée, des méthodes de prétraitement, des variables explicatives et des algorithmes utilisés pour améliorer la précision des prévisions. Cette section présente une synthèse des principaux travaux de recherche réalisés dans le domaine de la prévision de la consommation électrique et énergétique. L'objectif est d'identifier les méthodes les plus utilisées, d'analyser leurs performances et de mettre en évidence les limites et perspectives de ces approches afin de positionner le présent travail par rapport aux études existantes.

2.2 Les études précédentes

D'abord, Chou et Tran ont réalisé une étude sur la prévision de la consommation énergétique des bâtiments à partir des séries temporelles en utilisant des techniques de machine learning basées sur les habitudes d'utilisation des occupants résidentiels. Le but de cette étude était d'améliorer la précision des prévisions de consommation électrique en exploitant des données réelles et des modèles capables de représenter les comportements de consommation. Les auteurs ont comparé plusieurs modèles simples, modèles d'ensemble et modèles hybrides. Les résultats obtenus montrent que les modèles hybrides offrent généralement une meilleure

précision que les modèles simples, car ils combinent les avantages de plusieurs techniques de prévision. Cette étude conclut que l'utilisation du machine learning peut améliorer la gestion énergétique résidentielle, réduire les coûts et limiter l'impact environnemental. [43]

De plus, Deb et al. ont présenté une revue bibliographique sur les techniques de prévision des séries temporelles appliquées à la consommation énergétique des bâtiments. Les auteurs ont étudié plusieurs méthodes, notamment les modèles statistiques, les réseaux de neurones, les machines à vecteurs de support et les modèles hybrides. Leur objectif était d'identifier les approches les plus adaptées à la prévision énergétique dans les bâtiments existants disposant de données historiques. Les résultats montrent que les méthodes statistiques et les techniques de machine learning peuvent fournir des prévisions rapides et précises lorsque les données de consommation sont bien enregistrées. Les auteurs concluent que la qualité des données, le choix des variables et l'horizon de prévision jouent un rôle essentiel dans la performance des modèles. [44]

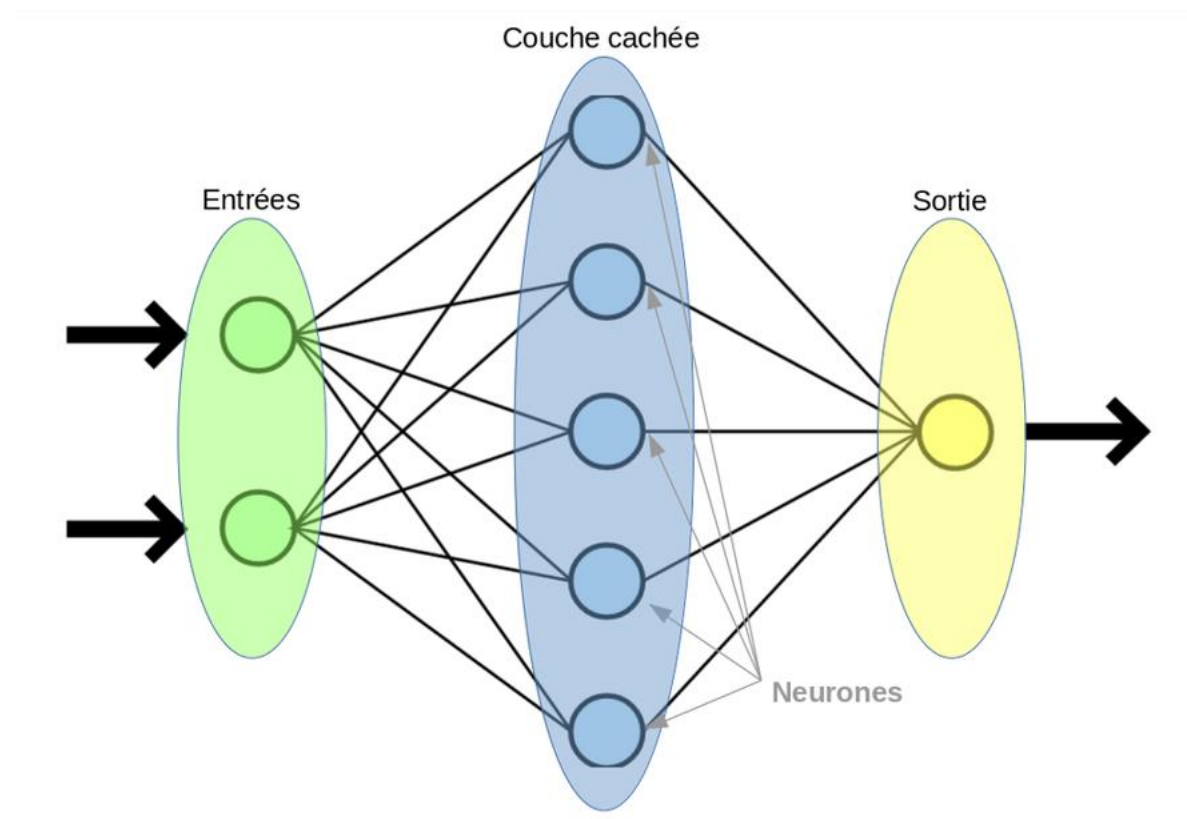


Figure 8: Architecture classique d'un réseau de neurones artificiels [45].

En addition, Amasyali et El-Gohary ont réalisé une revue approfondie des études basées sur les données pour la prédiction de la consommation énergétique des bâtiments. Le but de cette étude était d'analyser les types de bâtiments étudiés, les données utilisées, les algorithmes de machine learning appliqués et les mesures de performance adoptées. Les auteurs ont étudié des modèles tels que les réseaux de neurones artificiels, les machines à vecteurs de support, les arbres de décision et d'autres méthodes statistiques. Les résultats ont montré que les modèles pilotés par les données sont très utiles pour représenter les relations complexes entre les paramètres d'entrée et la consommation énergétique. Ils ont conclu que les futures recherches doivent accorder plus d'importance au prétraitement des données, au choix des variables et à la généralisation des modèles.[46]

Ensuite, Bourdeau et al. ont mené une étude de synthèse sur la modélisation et la prévision de la consommation énergétique des bâtiments à l'aide de techniques basées sur les données. Leur travail s'est concentré sur les caractéristiques des données d'entrée, les méthodes de prétraitement, les types de bâtiments, les horizons de prévision et les indicateurs d'évaluation. Les auteurs ont montré que les méthodes de machine learning sont largement utilisées grâce à leur capacité à apprendre les comportements complexes de consommation. Ils ont conclu que la performance des modèles dépend fortement de la préparation des données, de la sélection des variables explicatives et du choix de l'algorithme de prévision. [47]

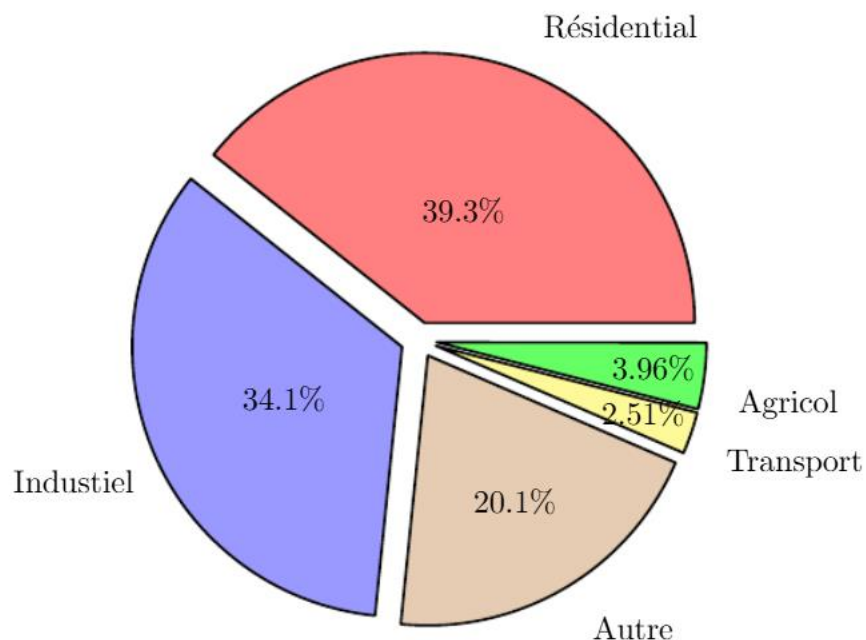


Figure 9 : Consommation d'électricité par secteur en Algérie 2021 (APRUE 2022)[48]

Puis, Nti et al. ont réalisé une revue systématique sur la prévision de la charge électrique. Le but de cette étude était d'analyser les principaux modèles utilisés dans le domaine de la prévision de la demande électrique. Les auteurs ont examiné plusieurs techniques, notamment les modèles statistiques, les réseaux de neurones artificiels, les machines à vecteurs de support et d'autres méthodes d'intelligence artificielle. Les résultats de cette revue montrent qu'une grande partie des modèles les plus utilisés repose sur l'intelligence artificielle, en particulier pour les prévisions à court terme où les profils de consommation sont complexes. Les auteurs concluent que les techniques d'IA sont devenues essentielles pour améliorer la précision des prévisions et assurer une meilleure gestion des réseaux électriques. [49]

Par ailleurs, Kuster et al. ont présenté une revue critique des modèles de prévision de la charge électrique. Le but de cette étude était d'aider les chercheurs et les utilisateurs à choisir le modèle le plus adapté selon le cas étudié, l'horizon de prévision et la disponibilité des données. Les auteurs ont analysé plusieurs méthodes de prévision utilisées dans les smart grids et les systèmes énergétiques modernes. Les résultats indiquent qu'il n'existe pas un modèle universel valable pour toutes les situations, car le choix du modèle dépend du type de données, de la fréquence d'échantillonnage et du niveau de précision recherché. Les auteurs concluent que la comparaison entre plusieurs modèles reste nécessaire avant de choisir la meilleure approche de prévision. [50]

De plus, Sun et al. ont réalisé une revue de l'état de l'art sur les approches pilotées par les données pour la prédiction énergétique des bâtiments. L'objectif principal était de présenter les étapes nécessaires au développement d'un modèle de prévision, notamment l'ingénierie des variables, le choix des modèles et l'interprétation des résultats. Les auteurs ont analysé 105 articles portant sur la prédiction énergétique des bâtiments. Les résultats montrent que les approches basées sur les données sont devenues très populaires, car elles permettent d'extraire automatiquement les relations statistiques à partir des données historiques. Les auteurs concluent que les modèles hybrides, combinant connaissances physiques et apprentissage automatique, peuvent améliorer la robustesse des prévisions.[51]

En outre, Amber et al. ont étudié la prévision de la consommation électrique d'un bâtiment administratif à Londres en comparant plusieurs techniques intelligentes. Les auteurs ont utilisé cinq méthodes : la régression multiple, la programmation génétique, les réseaux de neurones artificiels, les réseaux de neurones profonds et les machines à vecteurs de support. Le but de cette étude était d'évaluer la capacité de ces modèles à prévoir la consommation

électrique d'un bâtiment réel. Les résultats montrent que les techniques intelligentes peuvent fournir des prévisions fiables, avec une bonne capacité à modéliser les comportements non linéaires. Les auteurs concluent que les réseaux de neurones et les modèles avancés sont particulièrement utiles pour améliorer la précision de la prévision énergétique .[52]

Quant à Gellert et al., ils ont réalisé une étude sur la prévision de la consommation et de la production d'électricité dans les maisons intelligentes équipées de panneaux photovoltaïques. Les auteurs ont utilisé des méthodes statistiques telles que ARIMA et TBATS à partir de données issues d'un système de gestion énergétique domestique. L'objectif était de prévoir à la fois la consommation électrique et la production photovoltaïque dans une maison intelligente. Les résultats montrent que les méthodes statistiques peuvent donner de bonnes performances, notamment lorsque les données présentent des comportements saisonniers. Les auteurs concluent que la prévision énergétique dans les maisons intelligentes permet d'améliorer la gestion de l'électricité, surtout lorsque la maison combine consommation et production renouvelable. [53]

Ainsi, Bilal et al. ont mené une étude comparative sur la prévision de la consommation électrique des ménages à l'aide de modèles de séries temporelles et de machine learning. Ils ont utilisé plusieurs méthodes, notamment la régression linéaire, KNN, SVR, MLP, les processus gaussiens, ARIMA et VAR. Les données utilisées comprenaient des séries horaires et journalières de consommation électrique domestique. Les résultats montrent que SVR, MLP et les processus gaussiens donnent de bonnes performances pour la prévision de la consommation électrique des ménages. Les auteurs concluent que les modèles de machine learning sont adaptés à la prévision résidentielle, surtout lorsque les données contiennent des variations non linéaires et des informations météorologiques.[54]

De même, Peplinski et al. ont développé un cadre de machine learning pour estimer la demande électrique résidentielle à différentes échelles temporelles et spatiales. Les auteurs ont utilisé des données horaires de compteurs intelligents provenant d'environ 58 000 logements, ainsi que des données climatiques, des caractéristiques des bâtiments et des indicateurs socio-économiques. Le but de l'étude était de construire un modèle capable de prévoir la demande électrique résidentielle dans différents contextes. Les résultats montrent que l'intégration de plusieurs types de données améliore la précision de la prévision. Les auteurs concluent que les compteurs intelligents et les données environnementales sont très importants pour développer des modèles fiables de prévision résidentielle. [55]

En addition, *Ramnath et al.* ont étudié la prévision de la consommation électrique domestique en utilisant une approche hybride combinant le clustering flou et le machine learning. L'objectif de cette étude était d'améliorer la précision de la prédiction en regroupant les profils de consommation similaires avant l'application des modèles de prévision. Les auteurs ont montré que l'utilisation du clustering permet de mieux représenter les différents comportements de consommation des ménages. Les résultats indiquent que l'approche hybride peut améliorer les performances par rapport aux modèles appliqués directement sur l'ensemble des données. Les auteurs concluent que la combinaison entre classification des profils et modèles prédictifs constitue une solution efficace pour la prévision de la consommation domestique. [56]

Puis, *Jain et al.* ont évalué la prévision de la charge électrique à l'aide de plusieurs méthodes d'intelligence artificielle, notamment les machines à vecteurs de support, les réseaux LSTM, les modèles d'ensemble, les réseaux récurrents et les méthodes de deep learning. Le but de cette étude était de comparer les performances des modèles dans la prévision de la charge électrique à court terme. Les auteurs ont utilisé des données réelles d'une compagnie d'électricité sur plusieurs années. Les résultats montrent que les méthodes d'intelligence artificielle peuvent améliorer la précision des prévisions, en particulier lorsque les modèles sont adaptés aux caractéristiques temporelles de la demande. Les auteurs concluent que les modèles de machine learning et de deep learning sont prometteurs pour la gestion moderne de la charge électrique.[57]

Par la suite, *Baek et Seo* ont réalisé une étude sur la prévision de la demande d'électricité d'un campus universitaire en combinant des modèles de séries temporelles et de deep learning. Les auteurs ont utilisé SARIMA, SARIMAX, LSTM et un modèle hybride SARIMAX–LSTM. Ils ont également intégré des variables exogènes telles que les données météorologiques et les informations liées au calendrier académique. Les résultats montrent que le modèle hybride SARIMAX–LSTM donne de meilleures performances que les modèles individuels, avec un MAPE inférieur à 9 %. Les auteurs concluent que la combinaison entre les modèles statistiques et les réseaux de neurones permet de mieux capturer à la fois les relations linéaires, les saisonnalités et les comportements non linéaires de la demande électrique .[58]

Enfin, *Muqtadir et al.* ont proposé une méthode de prévision de charge résidentielle à très court terme basée sur des modèles de boosting, notamment LightGBM, XGBoost et CatBoost. Le but de cette étude était d'améliorer la prévision de la demande électrique

résidentielle en exploitant les avantages complémentaires des modèles d'ensemble. Les auteurs ont montré que l'intégration de plusieurs modèles de boosting permet de mieux gérer les données multivariées et de réduire les erreurs de prévision. Les résultats indiquent que cette approche est efficace pour les prévisions à très court terme, qui sont importantes pour les maisons intelligentes, la gestion de la charge et les réseaux électriques modernes. Les auteurs concluent que les modèles d'ensemble constituent une solution robuste pour la prévision de la demande électrique résidentielle .[59]

Chou et al. [60] ont réalisé une étude sur la prévision de la consommation énergétique des bâtiments à partir des séries temporelles. Ils ont utilisé des méthodes statistiques et des techniques de machine learning, ainsi que des modèles simples, des modèles d'ensemble et des modèles hybrides combinant prévision et optimisation. Ils ont obtenu que le modèle hybride offre une meilleure précision que les modèles simples et d'ensemble, ce qui permet d'améliorer la gestion énergétique et de réduire les coûts ainsi que l'impact environnemental.

Hossain et al. ont réalisé une étude sur la prévision de la demande énergétique dans les réseaux intelligents (smart grids) en utilisant l'analyse des séries temporelles. Ils ont utilisé des algorithmes de machine learning tels que KNN, SVM et la régression logistique, appliqués à des données de consommation énergétique collectées toutes les 15 minutes sur une période de deux ans. Ils ont évalué les performances à l'aide de métriques comme l'Accuracy, le Precision, le Recall et le F1-Score. Ils ont obtenu que les modèles KNN et SVM sont plus performants que la régression logistique, offrant une meilleure précision dans la prévision de la demande, ce qui permet d'améliorer l'efficacité des smart grids, la gestion de charge et la réduction des coûts énergétiques [61].

Baek et al. [62] ont réalisé une étude sur la prévision de la demande d'électricité d'une université en utilisant des séries temporelles et le deep learning. Ils ont utilisé des modèles SARIMA, SARIMAX et un modèle hybride SARIMAX-LSTM, en intégrant des variables exogènes (conditions météorologiques et calendrier académique) ainsi que des techniques de sélection de variables (analyse de corrélation, causalité de Granger, importance Random Forest, RFE et LASSO). Ils ont obtenu que le modèle hybride SARIMAX-LSTM est le plus performant avec un MAPE de 8,10 %, contre 15,26 % pour SARIMA, 10,62 % pour SARIMAX et 8,43 % pour LSTM, tout en réduisant l'erreur de 23,7 % (MAPE), 23,3 % (MAE) et 22,7 % (RMSE) par rapport aux meilleurs modèles individuels, ce qui améliore

significativement la précision des prévisions et la planification énergétique. La figure XX présente l'interface de prédiction utilisée pour ces auteurs.

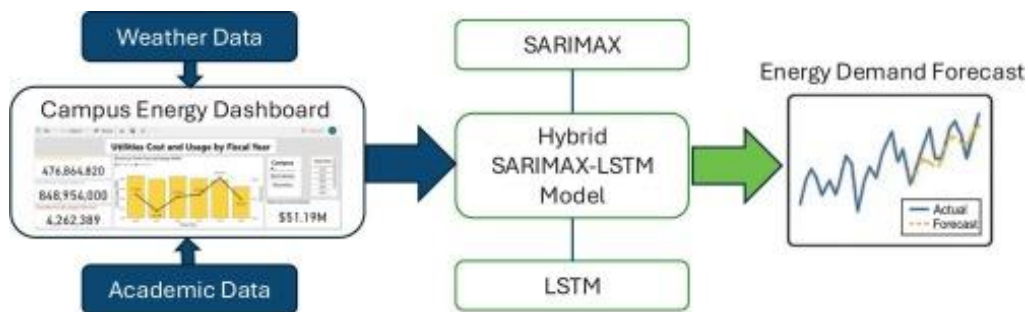


Figure 10:Schéma du modèle hybride SARIMAX-LSTM pour la prévision de la demande énergétique d'un campus universitaire [63]

Gellert et al. ont réalisé une étude sur la prévision de la consommation et de la production d'électricité dans les maisons intelligentes équipées de panneaux photovoltaïques. Ils ont utilisé des méthodes statistiques telles que ARIMA et TBATS, en les comparant à d'autres modèles, y compris des approches basées sur les réseaux de neurones, à partir de données issues d'un système de gestion énergétique domestique. Ils ont évalué les performances des modèles à l'aide de l'erreur absolue moyenne (MAE). Ils ont obtenu que le modèle TBATS est le plus performant avec une MAE de 73,62 Watts, surpassant les méthodes de prévision neuronales, ce qui permet une meilleure précision dans la prévision de la consommation et de la production énergétique. [64]

Yilu Wu et al. ont réalisé une étude sur l'application de la *régression linéaire* dans le domaine du Machine Learning, en mettant l'accent sur son rôle dans l'apprentissage supervisé. Ils ont utilisé le modèle de régression linéaire pour analyser et quantifier la relation entre une variable dépendante et une ou plusieurs variables explicatives à partir de données historiques, en s'appuyant sur des méthodes statistiques pour estimer les paramètres du modèle. L'étude a également mis en évidence les étapes de construction du modèle, notamment la collecte des données, l'entraînement, l'ajustement des coefficients et l'évaluation des performances. Ils ont obtenu que la régression linéaire est une méthode simple, rapide et facilement interprétable, capable de fournir des prédictions fiables lorsque les relations entre les variables sont linéaires, ce qui en fait un outil fondamental en Machine Learning. Cependant, ils ont également montré que ce modèle présente des limites lorsqu'il s'agit de modéliser des relations non linéaires complexes ou des données fortement bruitées. [65]

Jui-Sheng Chou **et al.** ont réalisé une étude sur la prévision de la consommation énergétique des bâtiments à partir des séries temporelles. Ils ont utilisé des méthodes statistiques ainsi que plusieurs techniques de Machine Learning et d'intelligence artificielle appliquées à des données réelles collectées en temps réel à partir d'un smart grid installé dans un bâtiment expérimental. Les auteurs ont analysé des modèles simples, des modèles d'ensemble et des modèles hybrides combinant des techniques de prévision et d'optimisation. Ils ont obtenu que le modèle hybride offre une précision supérieure aux modèles simples et aux modèles d'ensemble, démontrant ainsi une meilleure efficacité pour la prévision de la consommation énergétique et la planification de la gestion de l'énergie, tout en contribuant à la réduction des coûts énergétiques et de l'impact environnemental[66]

Hossain **et al.** ont réalisé une étude sur la prévision de la demande énergétique dans les smart grids en utilisant des techniques de Machine Learning appliquées aux séries temporelles. Ils ont utilisé des données de consommation énergétique résidentielles et commerciales collectées toutes les 15 minutes pendant une période de deux ans, permettant d'analyser les cycles horaires, les variations journalières et les effets saisonniers. Les auteurs ont appliqué trois modèles de classification, notamment KNN, SVM et la régression logistique, avec une séparation temporelle des données en ensembles d'entraînement et de test afin d'assurer la robustesse des modèles. Les performances ont été évaluées à l'aide des métriques Accuracy, Precision, Recall, F1-Score ainsi qu'une matrice de confusion. Ils ont obtenu que les modèles KNN et SVM présentent de meilleures performances que la régression logistique, avec un excellent équilibre entre précision et rappel, ce qui améliore l'efficacité opérationnelle des smart grids, la gestion des charges électriques et l'optimisation des coûts énergétiques.[67]

Jatin Bedi **et al.** ont réalisé une étude sur la prévision de la demande électrique en utilisant une approche basée sur le deep learning afin de mieux modéliser les dépendances à long terme dans les séries temporelles. Ils ont utilisé une méthode combinant l'analyse de clusters, la caractérisation des tendances de charge et des modèles LSTM (Long Short-Term Memory) multi-entrées multi-sorties pour prédire la consommation électrique selon les saisons, les jours et les intervalles de temps. Les auteurs ont également intégré une approche d'apprentissage actif basée sur une fenêtre glissante (moving window active learning) afin d'améliorer les performances de prédiction. L'approche proposée a été appliquée aux données de consommation électrique de Chandigarh en Inde et comparée à plusieurs modèles tels que les Artificial Neural Networks (ANN), les Recurrent Neural Networks (RNN) et la Support Vector Regression (SVR). Ils ont obtenu que le modèle basé sur les LSTM offre les

meilleures performances de prévision en capturant efficacement les dépendances temporelles à long terme et les comportements non linéaires de la consommation électrique, surpassant les modèles ANN, RNN et SVR.[68]

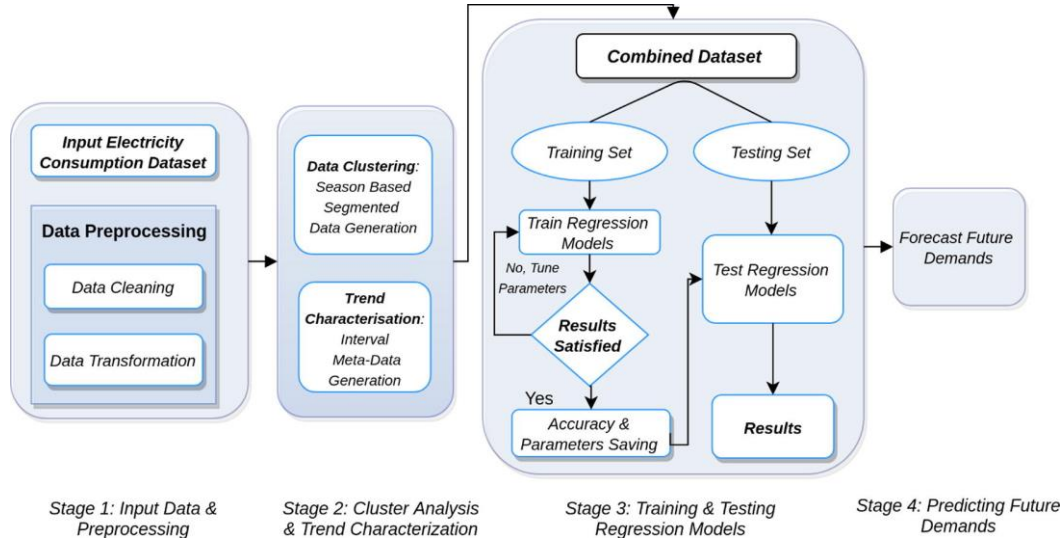


Figure 11: Architecture du Framework de prévision de la consommation électrique [68]

2.3 Conclusion

Les études précédentes montrent que la prévision de la consommation électrique a fait l'objet de nombreuses recherches utilisant des méthodes statistiques, de Machine Learning et de Deep Learning. Les résultats obtenus indiquent que les modèles avancés, notamment les approches hybrides, les réseaux LSTM et les méthodes de boosting telles que XGBoost et CatBoost, offrent généralement de meilleures performances que les méthodes traditionnelles. Ces travaux soulignent également l'importance de la qualité des données, du choix des variables explicatives et des techniques de prétraitement pour améliorer la précision des prévisions énergétiques.

Chapitre III : Méthodologie de la recherche

Dans ce chapitre, nous décrivons chaque paramètre d'entrée et de sortie, en soulignant son rôle et son importance dans notre étude. Nous avons ensuite abordé la sélection des modèles de machine learning, en justifiant nos choix selon les critères de performance, de complexité et d'ajustement avec les données disponibles. Enfin, nous identifions les défis et les considérations rencontrées, tels que la qualité des données, le surapprentissage ou les limitations de calcul, tout en proposant des moyens pour les surmonter. Cette analyse approfondie fournit une base méthodologique solide pour nos travaux ultérieurs

3.1 Introduction

Après avoir étudié les travaux de recherche antérieurs relatifs à la prévision de la consommation énergétique, il est nécessaire de présenter la méthodologie adoptée dans le cadre de cette étude. Ce chapitre décrit les différentes étapes suivies pour la mise en œuvre des modèles de prédiction. Dans un premier temps, les données utilisées ainsi que les variables d'entrée et de sortie sont présentées. Ensuite, les différentes étapes de prétraitement des données sont détaillées afin d'assurer leur qualité et leur exploitation optimale. Enfin, les modèles de Machine Learning retenus, à savoir la Régression Linéaire, Random Forest, XGBoost et CatBoost, sont décrits, ainsi que les métriques utilisées pour évaluer leurs performances.

3.2 Description des paramètres d'entrée et de sortie

Dans le cadre de cette étude, les données utilisées proviennent de la plateforme Kaggle, une source reconnue pour la mise à disposition de jeux de données utilisés en analyse de données et en apprentissage automatique. Plus précisément, le dataset exploité est intitulé *Household Energy Consumption*, qui regroupe des informations relatives à la consommation énergétique d'un foyer.[42]

Ces données représentent la consommation électrique domestique sur une période donnée, enregistré à Avril 2025 (90,000 Records)

. Elles permettent d'analyser le comportement énergétique des utilisateurs et d'identifier les facteurs influençant la consommation.[69]

Le jeu de données comprend plusieurs variables décrivant à la fois la consommation énergétique et les conditions associées. Parmi les principales variables utilisées dans cette étude, on retrouve :

- **La consommation électrique (kWh)** : il s'agit de la variable cible à prédire, représentant la quantité d'énergie consommée.
- **La température** : une variable explicative importante influençant la consommation, notamment en raison de l'utilisation des systèmes de chauffage ou de climatisation.
- **Le temps (date et heure)** : cette variable permet de capturer les effets temporels tels que les variations journalières, hebdomadaires et saisonnières.

- **Autres variables potentielles** : selon le dataset, d'autres facteurs comme l'humidité, la vitesse du vent ou des indicateurs temporels (jour, mois, heure) peuvent être inclus afin d'améliorer la qualité des prévisions.

Tableau 1 - Household Energy Consumption - April 2025 (90,000 Records) [70]

Column Name	Data Type Category	Description
Household_ID	Categorical (Nominal)	Unique identifier for each household
Date	Datetime	The date of the energy usage record
Energy_Consumption_kWh	Numerical (Continuous)	Total energy consumed by the household in kWh
Household_Size	Numerical (Discrete)	Number of individuals living in the household
Avg_Temperature_C	Numerical (Continuous)	Average daily temperature in degrees Celsius
Has_AC	Categorical (Binary)	Indicates if the household has air conditioning (Yes/No)
Peak_Hours_Usage_kWh	Numerical (Continuous)	Energy consumed during peak hours in kWh

L'objectif principal de cette étude est de **prévoir la consommation électrique** à partir des variables explicatives disponibles, en utilisant des techniques de **Machine Learning**.

Pour ce faire, plusieurs modèles ont été implémentés à l'aide du langage de programmation **Python**, dans l'environnement **Anaconda**. Les modèles utilisés sont les suivants :

- **La régression linéaire**
- **Le Random Forest**
- **XGBoost**
- **CatBoost**

Ces modèles ont été choisis afin de comparer les performances entre une méthode simple et interprétable (régression linéaire) et des méthodes plus avancées capables de capturer des relations non linéaires complexes (Random Forest, XGBoost et CatBoost).[71]

Ainsi, ce jeu de données constitue une base pertinente pour l'analyse et la modélisation de la consommation énergétique résidentielle, permettant d'évaluer l'efficacité de différents modèles de prévision.

3.3 Présentation des différents modèles de prédiction utilisant l'IA

3.3.1 Régression linéaire:

La régression linéaire est un algorithme qui fournit une relation linéaire entre une variable indépendante et une variable dépendante afin de prédire l'issue d'événements futurs. Il s'agit d'une méthode statistique utilisée en science des données et en apprentissage automatique pour l'analyse prédictive. La variable indépendante est également la variable prédictive ou explicative qui reste inchangée en raison de la variation d'autres variables. En revanche, la variable dépendante évolue avec les fluctuations de la variable indépendante. Le modèle de régression prédit la valeur de la variable dépendante, qui est la variable de réponse ou de résultat analysée ou étudiée. Ainsi, la régression linéaire est un algorithme d'apprentissage supervisé qui simule une relation mathématique entre des variables et effectue des prédictions pour des variables continues ou numériques telles que les ventes, le salaire, l'âge, le prix des produits, etc. la Figure représente Régression linéaire[72] .

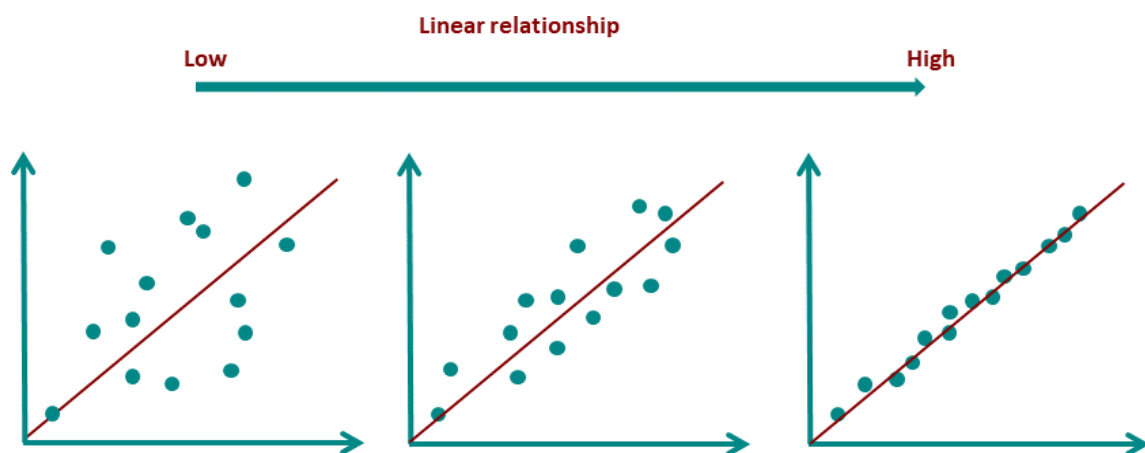


Figure 12: Régression linéaire [72]

Random Forest

Le modèle Random Forest est une méthode d'apprentissage supervisé basée sur un ensemble d'arbres de décision. Chaque arbre est construit à partir d'un échantillon aléatoire des données et tente de prédire la consommation électrique à partir des variables explicatives. La prédiction finale est obtenue par agrégation des résultats de tous les arbres, par exemple par moyenne pour les tâches de régression. Les principaux points forts de Random Forest sont : Capacité à capturer des relations non linéaires et complexes : contrairement à la régression linéaire, il peut modéliser des comportements irréguliers et des interactions multiples entre

variables. Robustesse au surapprentissage (overfitting) : grâce à l'utilisation de nombreux arbres et à l'agrégation des résultats, Random Forest réduit le risque que le modèle s'ajuste trop aux données d'entraînement. Traitement des données bruitées et manquantes : le modèle est tolérant aux erreurs dans les données ou aux valeurs manquantes, ce qui est fréquent dans les mesures réelles de consommation électrique[73]. Importance des variables : Random Forest fournit une mesure de l'importance de chaque variable dans la prédiction, ce qui permet de mieux comprendre quels facteurs influencent le plus la demande énergétique. En pratique, Random Forest est particulièrement adapté à la prévision énergétique lorsque la consommation dépend de multiples facteurs interconnectés et non linéaires. Il est largement utilisé dans les bâtiments intelligents, les smart grids et pour la gestion de l'énergie résidentielle et industrielle. Cependant, il nécessite un temps de calcul plus long que la régression linéaire et peut être moins interprétable à cause de sa structure en ensemble d'arbres multiples.

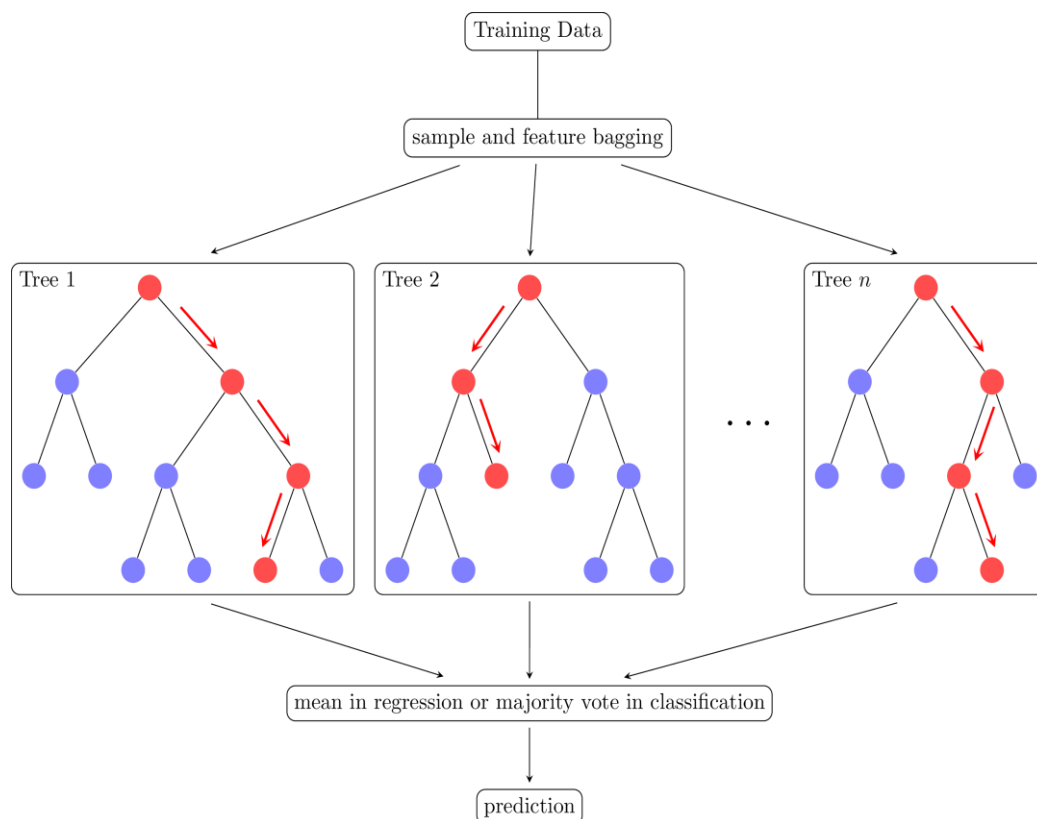


Figure 13: Random forest [74]

3.3.2 Modèle XGBoost

Le modèle **XGBoost** (Extreme Gradient Boosting) est un algorithme d'apprentissage supervisé basé sur la technique du boosting. Il consiste à construire une succession d'arbres de décision, où chaque nouvel arbre est entraîné pour corriger les erreurs commises par les modèles précédents. Contrairement aux méthodes classiques, XGBoost introduit des mécanismes avancés tels que la régularisation, permettant de réduire le risque de surapprentissage et d'améliorer la généralisation du modèle[75]. Il utilise également une optimisation basée sur le gradient, ce qui rend l'apprentissage plus rapide et plus efficace.

Dans le cadre de la prévision de la demande électrique, XGBoost est particulièrement performant grâce à sa capacité à modéliser des relations non linéaires entre les variables, telles que la température, l'heure ou les habitudes de consommation. Il peut également gérer efficacement les données manquantes et les grands volumes de données. Cependant, ce modèle nécessite un réglage précis de plusieurs hyperparamètres, tels que la profondeur des arbres, le taux d'apprentissage et le nombre d'itérations, afin d'obtenir des résultats optimaux.

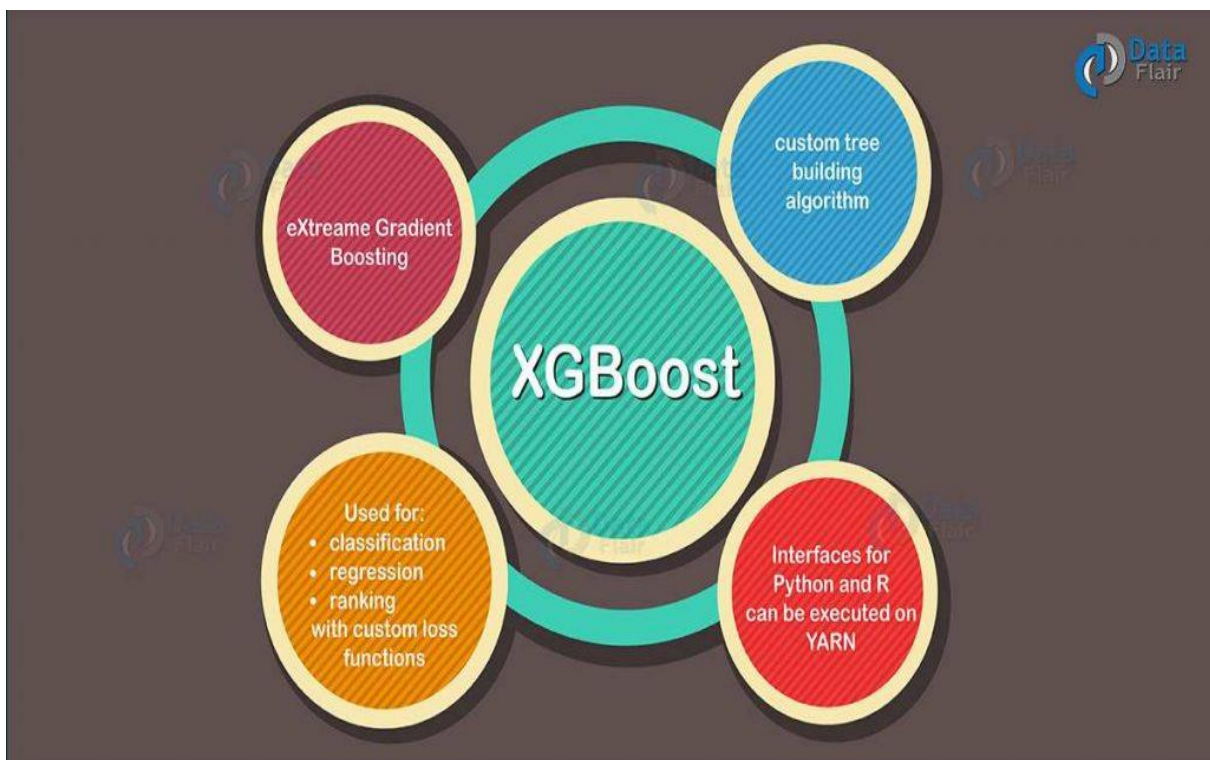


Figure 14 : XGBoost Algorithm – Objective [76]

3.3.3 Modèle CatBoost

Le modèle **CatBoost** (Categorical Boosting) est une méthode d'apprentissage automatique basée sur le boosting, développée pour améliorer le traitement des variables catégorielles. Contrairement à de nombreux algorithmes qui nécessitent une transformation préalable des variables qualitatives, CatBoost intègre directement ces variables dans le processus d'apprentissage. Cela permet de simplifier le prétraitement des données tout en améliorant la précision du modèle. CatBoost utilise également des techniques avancées pour réduire le surapprentissage, notamment un ordre spécifique d'apprentissage et une gestion optimisée des données[77]. Ces caractéristiques le rendent particulièrement adapté aux problèmes de prévision où les données comportent des variables comme les jours de la semaine, les saisons ou les types d'utilisateurs. Dans le contexte de la consommation électrique résidentielle, CatBoost offre de bonnes performances en capturant les interactions complexes entre les différentes variables explicatives. Il est également reconnu pour sa stabilité et sa robustesse. Néanmoins, comme XGBoost, ce modèle nécessite un ajustement des hyperparamètres et peut être plus complexe à interpréter que les modèles statistiques traditionnels.[78]

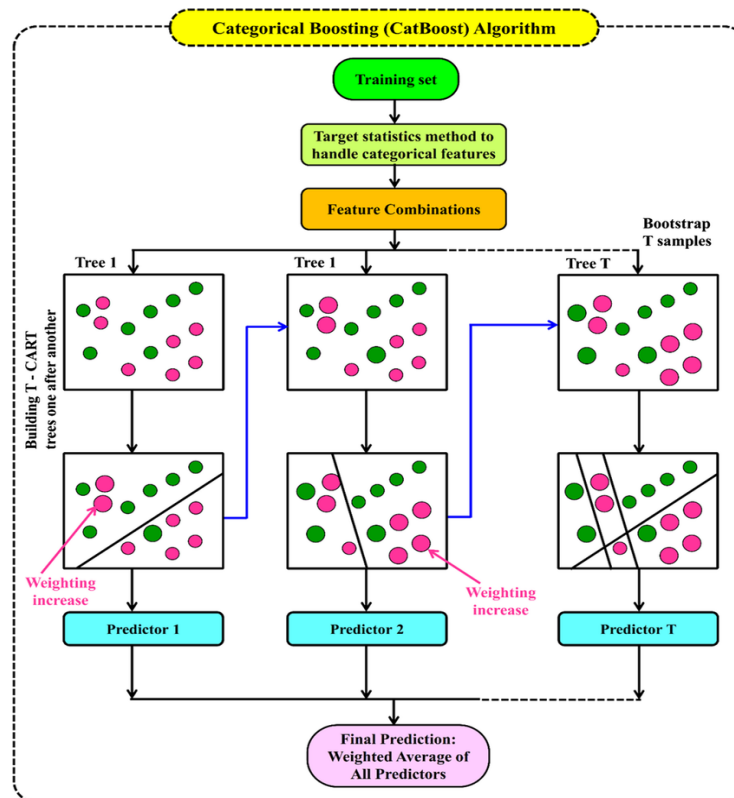


Figure 15 :The flow diagram of the CatBoost model[77]

3.4 Prétraitement des données

Le prétraitement des données constitue une étape essentielle dans le processus de modélisation car il permet d'améliorer la qualité des données et d'optimiser les performances des modèles de **Machine Learning**. Les données brutes issues du dataset Household Energy Consumption ont subi plusieurs étapes de transformation avant leur utilisation.[79]

3.5 Nettoyage des données

Dans un premier temps, une phase de nettoyage a été réalisée afin de traiter les valeurs manquantes, les incohérences et les éventuelles anomalies présentes dans le jeu de données. Les lignes contenant des valeurs nulles ou aberrantes ont été soit supprimées, soit corrigées selon leur impact sur l'analyse. Cette étape permet de garantir la fiabilité des résultats obtenus.[80]

3.6 Transformation des variables temporelles

Les variables liées au temps (date et heure) ont été transformées en variables exploitables par les modèles. Ainsi, plusieurs attributs ont été extraits, tels que :

- l'heure de la journée,
- le jour,
- le mois,
- éventuellement le jour de la semaine.

Cette transformation permet de capturer les comportements saisonniers et les variations périodiques de la consommation électrique.

3.7 Normalisation et mise à l'échelle

Afin d'assurer une meilleure convergence des modèles et d'éviter que certaines variables dominent les autres, les données ont été normalisées ou standardisées. Cette étape est particulièrement importante pour les modèles sensibles à l'échelle des données, comme la régression linéaire.

3.8 Séparation des données (Train/Test)

Le jeu de données a été divisé en deux parties :

- un ensemble d'entraînement (**training set**) utilisé pour construire les modèles,
- un ensemble de test (**test set**) utilisé pour évaluer leurs performances.

Cette séparation permet de vérifier la capacité des modèles à généraliser sur de nouvelles données.[81]

3.9 Sélection des variables

Les variables les plus pertinentes pour la prédiction de la consommation ont été sélectionnées. Cette étape permet de réduire la complexité du modèle et d'améliorer sa précision en éliminant les variables peu significatives.

3.10 Évaluation des performances des modèles

Afin de comparer les performances des différents modèles, plusieurs métriques d'évaluation ont été utilisées. Ces indicateurs permettent de mesurer la précision des prédictions en comparant les valeurs prédites aux valeurs réelles.

3.10.1 Erreur absolue moyenne (MAE)

La MAE (Mean Absolute Error) mesure la moyenne des erreurs absolues entre les valeurs réelles et les valeurs prédites. Elle donne une indication simple de l'erreur moyenne du modèle.

3.10.2 Erreur quadratique moyenne (RMSE)

La RMSE (Root Mean Square Error) mesure la racine carrée de la moyenne des carrés des erreurs. Elle pénalise davantage les grandes erreurs, ce qui la rend utile pour évaluer la robustesse du modèle.

3.10.3 Coefficient de détermination (R^2)

Le coefficient R^2 mesure la proportion de la variance expliquée par le modèle. Une valeur proche de 1 indique une bonne capacité de prédiction.

3.10.4 Analyse comparative des modèles

Les modèles développés ont été comparés en fonction de ces métriques afin d'identifier celui qui offre les meilleures performances. En général :

- la régression linéaire sert de référence,
- Random Forest améliore la capture des relations non linéaires,
- XGBoost et CatBoost offrent souvent les meilleures performances grâce à leurs mécanismes d'optimisation avancés.

3.11 Conclusion

Dans ce chapitre, nous avons présenté la méthodologie suivie pour la réalisation de cette étude. Après avoir décrit le jeu de données et les différentes variables considérées, plusieurs étapes de prétraitement ont été effectuées afin d'améliorer la qualité des données et de préparer leur utilisation dans les modèles de prédiction. Nous avons également présenté les différents modèles de Machine Learning retenus ainsi que les métriques d'évaluation utilisées pour comparer leurs performances. Cette démarche méthodologique constitue une base essentielle pour l'obtention de résultats fiables et précis. Le chapitre suivant sera consacré à la présentation, à l'analyse et à la discussion des résultats obtenus.

Chapitre IV: Résultats et discussions

Dans ce chapitre, nous présentons et analysons les résultats obtenus à partir des différents modèles de prévision de la consommation électrique développés dans le chapitre précédent. Après avoir décrit la méthodologie adoptée, le prétraitement des données ainsi que les modèles de machine learning utilisés, il est essentiel d'évaluer leurs performances afin de déterminer leur efficacité dans la prédiction de la demande énergétique.

Dans un premier temps, les résultats des prévisions seront exposés pour chaque modèle, en mettant en évidence leur capacité à reproduire le comportement réel de la consommation électrique. Ensuite, une comparaison sera effectuée entre les différents modèles, notamment la régression linéaire, Random Forest, XGBoost et CatBoost, afin d'identifier celui qui offre les meilleures performances selon les métriques d'évaluation utilisées.

Par la suite, une analyse approfondie des résultats permettra d'interpréter les écarts observés entre les valeurs réelles et les valeurs prédites, tout en mettant en évidence les forces et les limites de chaque approche. Enfin, ce chapitre abordera l'apport de la prévision dans la gestion énergétique d'une habitation, en montrant comment ces modèles peuvent contribuer à une meilleure optimisation de la consommation électrique.

4.1 Introduction

Dans ce chapitre, nous présentons et analysons les résultats obtenus à partir des différents modèles de prévision de la consommation électrique développés dans le chapitre précédent. Après avoir décrit la méthodologie adoptée, le prétraitement des données ainsi que les modèles de machine learning utilisés, il est essentiel d'évaluer leurs performances afin de déterminer leur efficacité dans la prédiction de la demande énergétique.

Dans un premier temps, les résultats des prévisions seront exposés pour chaque modèle, en mettant en évidence leur capacité à reproduire le comportement réel de la consommation électrique. Ensuite, une comparaison sera effectuée entre les différents modèles, notamment la régression linéaire, Random Forest, XGBoost et CatBoost, afin d'identifier celui qui offre les meilleures performances selon les métriques d'évaluation utilisées.

Par la suite, une analyse approfondie des résultats permettra d'interpréter les écarts observés entre les valeurs réelles et les valeurs prédites, tout en mettant en évidence les forces et les limites de chaque approche. Enfin, ce chapitre abordera l'apport de la prévision dans la gestion énergétique d'une habitation, en montrant comment ces modèles peuvent contribuer à une meilleure optimisation de la consommation électrique.

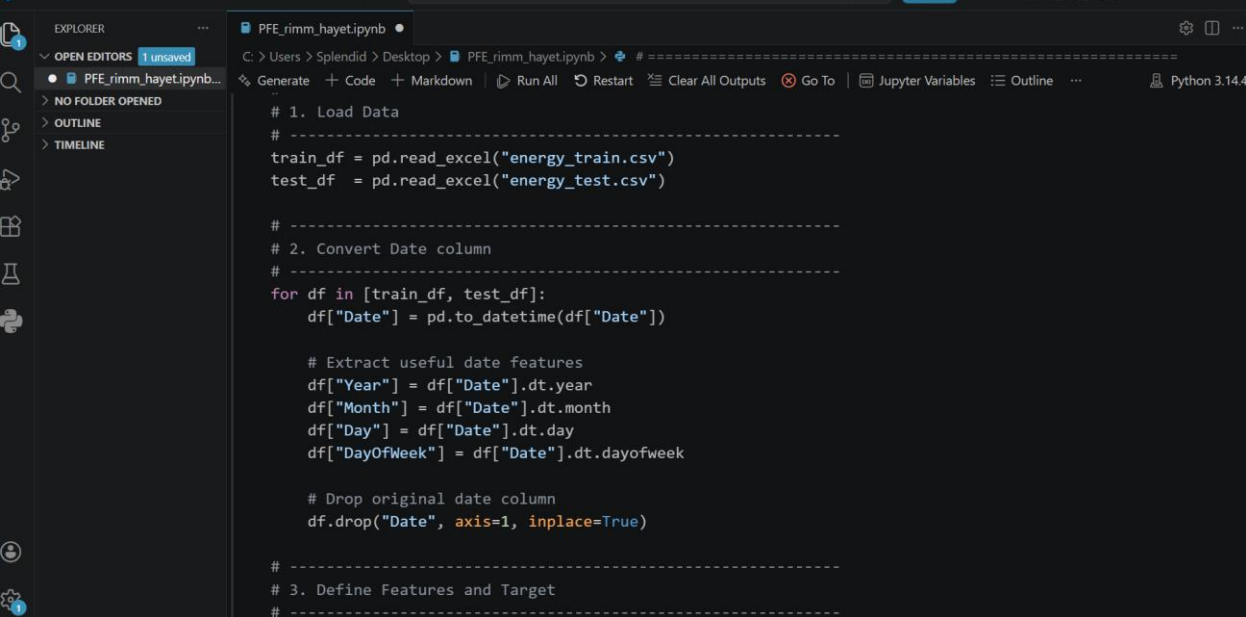
4.2 Présentation de Python

Python est un langage de programmation polyvalent utilisé pour l'analyse de données, le développement d'applications et la création de modèles prédictifs. Il permet aux utilisateurs de manipuler des données, d'effectuer des calculs complexes et de visualiser les résultats de manière efficace. Créé par Guido van Rossum, Python est aujourd'hui largement utilisé dans les domaines scientifiques et industriels.

Dans le cadre de ce travail, Python est utilisé pour la prévision de la consommation électrique, en exploitant des séries temporelles contenant des informations telles que la consommation, la température, le temps (heures, jours) et d'autres variables explicatives. Contrairement à des outils comme Orange3 qui reposent sur une approche visuelle, Python offre une approche basée sur la programmation, permettant un contrôle total sur chaque étape du processus analytique.

L'utilisation de Python permet de construire un pipeline complet de traitement des données, comprenant les étapes suivantes :

- **Préparation & Exploration** : Importation avec Pandas, nettoyage des données et analyse visuelle des tendances de consommation.
- **Modélisation** : Mise en œuvre d'algorithmes variés, allant de la régression classique aux modèles de boosting (XGBoost, CatBoost).
- **Évaluation** : Calcul de la précision via les métriques RMSE, MAE et R^2 pour sélectionner le modèle le plus performant.
- **Visualisation** : Comparaison graphique des valeurs réelles et prédites pour valider la qualité finale des prévisions.



```
EXPLORER
  OPEN EDITORS 1 unsaved
    PFE_rimm_hayetIpynb...
  NO FOLDER OPENED
  OUTLINE
  TIMELINE

PFE_rimm_hayetIpynb
C:\Users\Splendid\Desktop> PFE_rimm_hayetIpynb
# =====
# 1. Load Data
# -----
train_df = pd.read_excel("energy_train.csv")
test_df = pd.read_excel("energy_test.csv")

# -----
# 2. Convert Date column
# -----
for df in [train_df, test_df]:
    df["Date"] = pd.to_datetime(df["Date"])

    # Extract useful date features
    df["Year"] = df["Date"].dt.year
    df["Month"] = df["Date"].dt.month
    df["Day"] = df["Date"].dt.day
    df["DayOfWeek"] = df["Date"].dt.dayofweek

    # Drop original date column
    df.drop("Date", axis=1, inplace=True)

# -----
# 3. Define Features and Target
# -----
```

Figure 16 : Image réelle du programme de la simulation de notre étude

Métriques d'évaluation (R^2 , MAE, MAPE, RMSE, MSE) : Plusieurs métriques sont utilisées afin de mesurer la performance du modèle, notamment le coefficient de détermination (R^2), l'erreur absolue moyenne (MAE), l'erreur quadratique moyenne (MSE), ainsi que la racine de l'erreur quadratique moyenne (RMSE) et le MAPE.

Le coefficient R^2 permet d'évaluer la qualité d'ajustement du modèle : plus sa valeur est proche de 1, meilleure est l'explication de la variabilité des données par le modèle. En revanche, les métriques MAE, MSE, RMSE et MAPE doivent être les plus faibles possibles, car des valeurs réduites indiquent que les prédictions sont proches des valeurs réelles et que le modèle présente de bonnes performances.

R^2 représente la proportion de la variabilité de la variable cible expliquée par les variables explicatives, tandis que les autres indicateurs quantifient l'erreur de prédiction du modèle.

$$R^2 = 1 - \frac{\sum_{i=1}^n (\mathbf{y}_i - \hat{\mathbf{y}}_i)^2}{\sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})^2}$$

RMSE quantifie l'écart entre les valeurs prédites et les valeurs réelles en tenant compte de l'écart-type des erreurs.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}$$

MAE mesure la différence absolue entre les valeurs prédites et les valeurs réelles.

$$MAE = \frac{\sum_{t=1}^n |y_t - \hat{y}_t|}{n}$$

MSE (Mean Squared Error) correspond à l'erreur quadratique moyenne .

$$MSE = \frac{1}{n} \sum_{i=1}^n (\mathbf{y}_i - \hat{\mathbf{y}}_i)^2$$

où n représente le nombre d'échantillons, y_t désigne la valeur réelle de la variable cible, et $\hat{y}_t$ correspond à la valeur prédite.

4.3 Résultats de la prévision

La Figure 17 est un diagramme en violon illustrant la distribution de la variable Has_AC, qui indique la présence (1) ou l'absence (0) d'un système de climatisation dans les foyers. L'axe vertical comporte uniquement deux modalités, 0 et 1, correspondant à la nature binaire de la variable. La forme du violon reflète la densité de distribution des observations pour ces deux catégories. On observe une concentration des données exclusivement autour de ces deux valeurs, sans dispersion intermédiaire, ce qui confirme le caractère strictement discret de la variable. La largeur relative du violon au niveau de 0 et de 1 traduit la proportion de foyers sans climatisation et avec climatisation dans l'échantillon. Au centre du violon, un boxplot est superposé. Le point blanc représente la médiane, qui se situe nécessairement sur l'une des deux valeurs. Le rectangle noir (intervalle interquartile) est très réduit en raison de l'absence de variabilité continue. Dans l'ensemble, cette figure met en évidence une répartition binaire nette des foyers, permettant d'identifier rapidement la dominance éventuelle d'une catégorie.

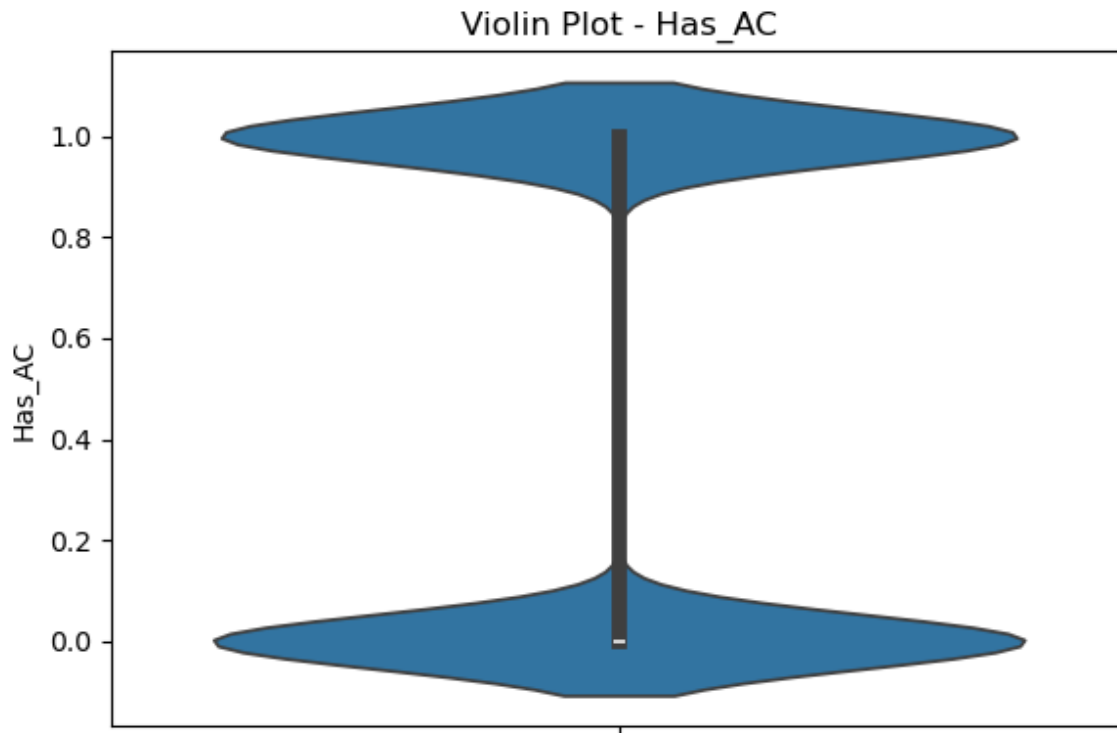


Figure 17 : Violin plot de Has_AC

La **Figure 18** présentée est un diagramme en violon illustrant la distribution de la variable Température moyenne (°C). L'axe vertical s'étend approximativement de 10 °C à 25 °C, couvrant l'ensemble des valeurs observées. La forme du violon reflète la densité de distribution des températures. On observe que la partie la plus large se situe autour de 16 à 19 °C, indiquant une forte concentration des valeurs dans cette plage. La distribution apparaît relativement symétrique, avec une légère dispersion vers les valeurs extrêmes. Les extrémités plus fines du violon traduisent la rareté des températures proches de 10 °C ou 25 °C. Cela montre que ces valeurs existent, mais restent peu fréquentes. Le boxplot central met en évidence la médiane (point blanc), ainsi que l'intervalle interquartile contenant les 50 % des données centrales. Les moustaches s'étendent vers les valeurs minimales et maximales non aberrantes. Globalement, cette figure met en évidence une distribution relativement stable de la température, avec une concentration autour de valeurs modérées et une variabilité limitée.

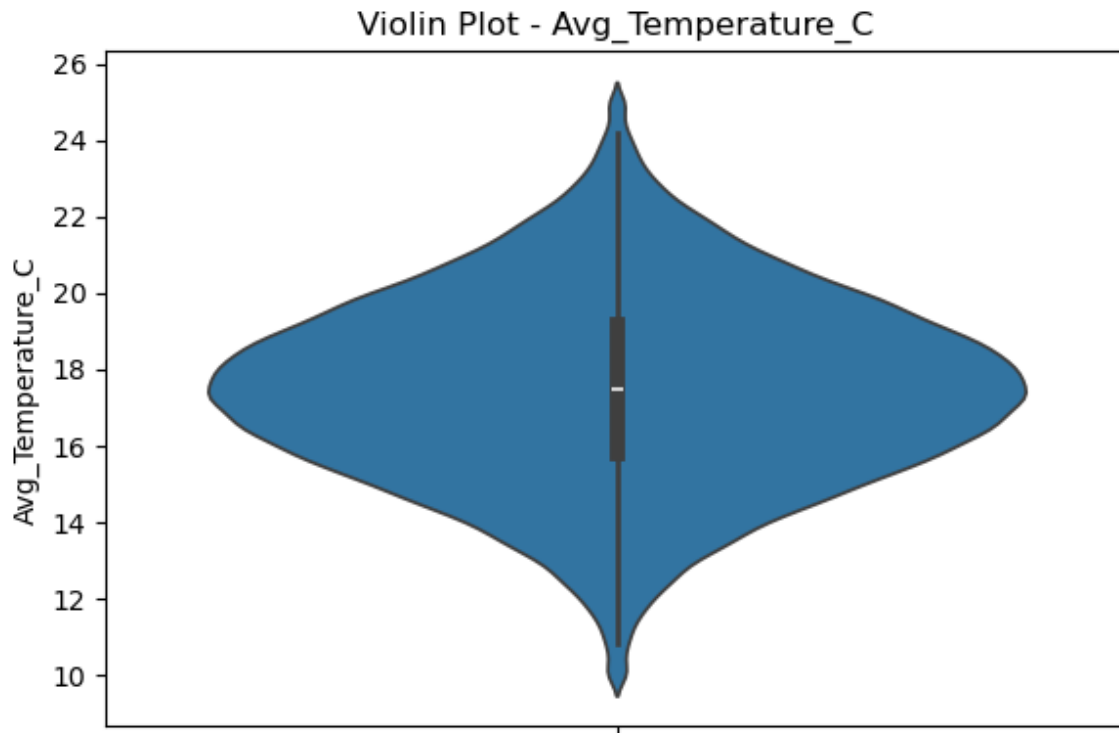


Figure 18 : Violin plot de Avg_Temperature_C

La figure présentée est un diagramme en violon illustrant la distribution de la variable Taille du foyer (nombre de personnes). L'axe vertical varie de 1 à 6 personnes, correspondant aux différentes tailles de foyers observées. La forme du violon montre une distribution discrète, avec des niveaux distincts correspondant à chaque valeur entière. La largeur du violon à chaque niveau indique la fréquence des foyers ayant cette taille. On observe que certaines tailles de foyers présentent une densité plus importante, traduisant une concentration plus élevée des observations. À l'inverse, les tailles extrêmes (très petits ou très grands foyers) sont généralement moins fréquentes. Le boxplot central indique la médiane ainsi que l'intervalle interquartile. Toutefois, en raison du caractère discret de la variable, ces indicateurs sont moins lisses que pour une variable continue.

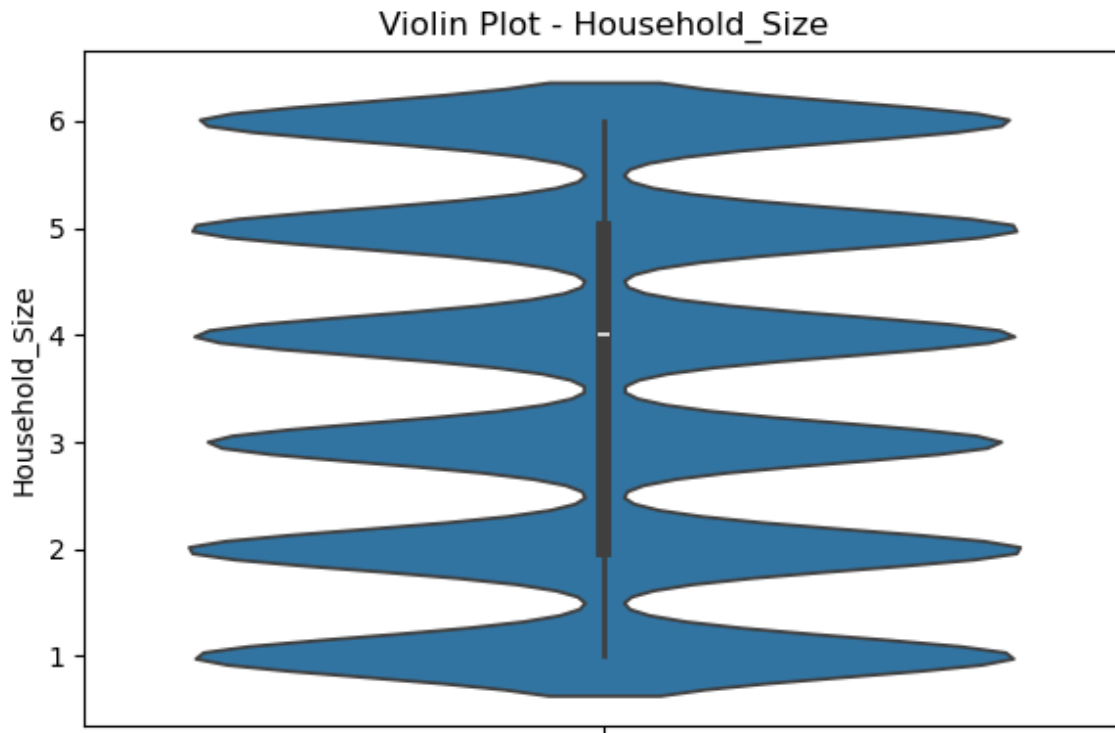


Figure 19 : Violin plot de Household_Size

La figure présentée est un ensemble de diagrammes en violon comparant la distribution de plusieurs variables numériques : Household_Size, Avg_Temperature_C, Peak_Hours_Usage_kWh et Energy_Consumption_kWh. Chaque violon représente la distribution d'une variable spécifique. On observe que la variable Peak_Hours_Usage_kWh présente une distribution étendue, avec des valeurs comprises approximativement entre 0 et 10 kWh, indiquant une variabilité importante de l'utilisation pendant les heures de pointe. De même, la variable Energy_Consumption_kWh montre une dispersion encore plus large, avec des valeurs allant d'environ 0 à 20 kWh, ce qui reflète une forte variabilité de la consommation énergétique totale. En comparaison, les variables Avg_Temperature_C (10 à 25 °C) et Household_Size (1 à 6 personnes) présentent des distributions plus resserrées, indiquant une variabilité plus limitée. La largeur des violons met en évidence les zones de forte concentration des données pour chaque variable, tandis que les boxplots centraux indiquent les médianes et les intervalles interquartiles. Dans l'ensemble, cette figure comparative montre clairement que les variables liées à la consommation (Peak_Hours_Usage_kWh et Energy_Consumption_kWh) sont les plus dispersées, ce qui confirme leur rôle central et leur forte variabilité dans l'analyse.

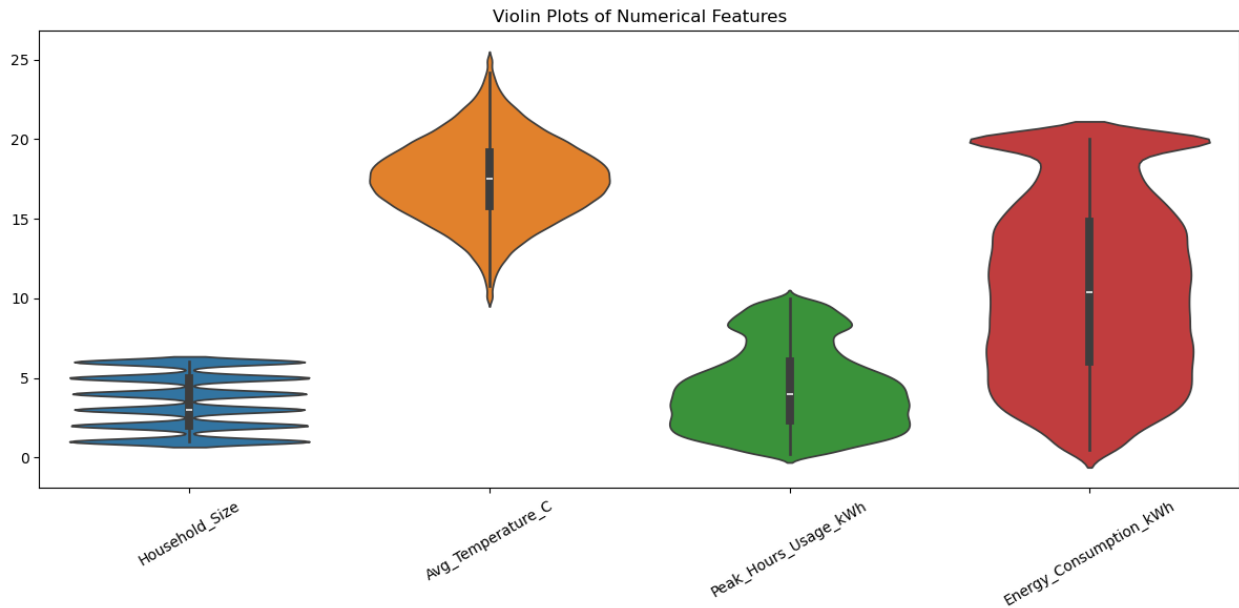


Figure 20 : Violin plots des variables numériques

La matrice de corrélation présente les relations linéaires entre les paramètres. Les valeurs proches de 1 indiquent une relation positive forte, alors que les valeurs proches de 0 indiquent une relation faible. On observe une très forte corrélation entre Peak_Hours_Usage_kWh et Energy_Consumption_kWh (0,97), une forte corrélation entre Household_Size et Energy_Consumption_kWh (0,89), ainsi qu'une relation importante entre Household_Size et Peak_Hours_Usage_kWh (0,79). Cela montre que la taille du foyer et l'utilisation pendant les heures de pointe sont des facteurs très influents dans la consommation énergétique.

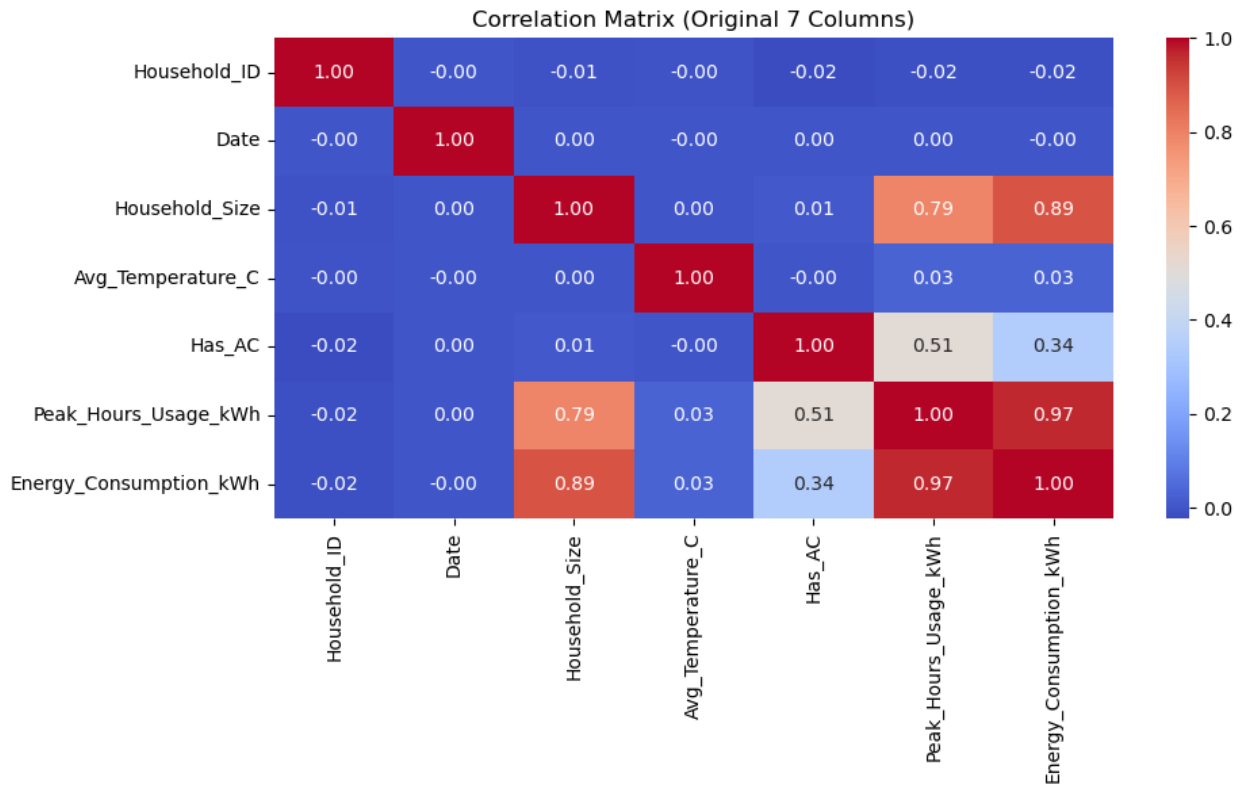


Figure 21: Matrice de corrélation des sept paramètres originaux

La Figure 22 présente la méthodologie adoptée pour la prévision de la demande électrique à l'aide de techniques d'apprentissage automatique. Après la collecte des données provenant de différentes sources (historique de consommation, variables météorologiques, informations temporelles et facteurs externes), celles-ci sont prétraitées afin d'assurer leur qualité et leur pertinence. Les données préparées sont ensuite utilisées pour entraîner quatre modèles de régression, à savoir XGBoost, CatBoost, la Régression Linéaire et Random Forest. Les performances des modèles sont comparées à l'aide des métriques MAE, RMSE, MAPE et R^2 , permettant d'identifier le modèle le plus performant. Les résultats obtenus contribuent à une prévision plus précise de la demande électrique, favorisant ainsi une meilleure planification et une prise de décision plus efficace..

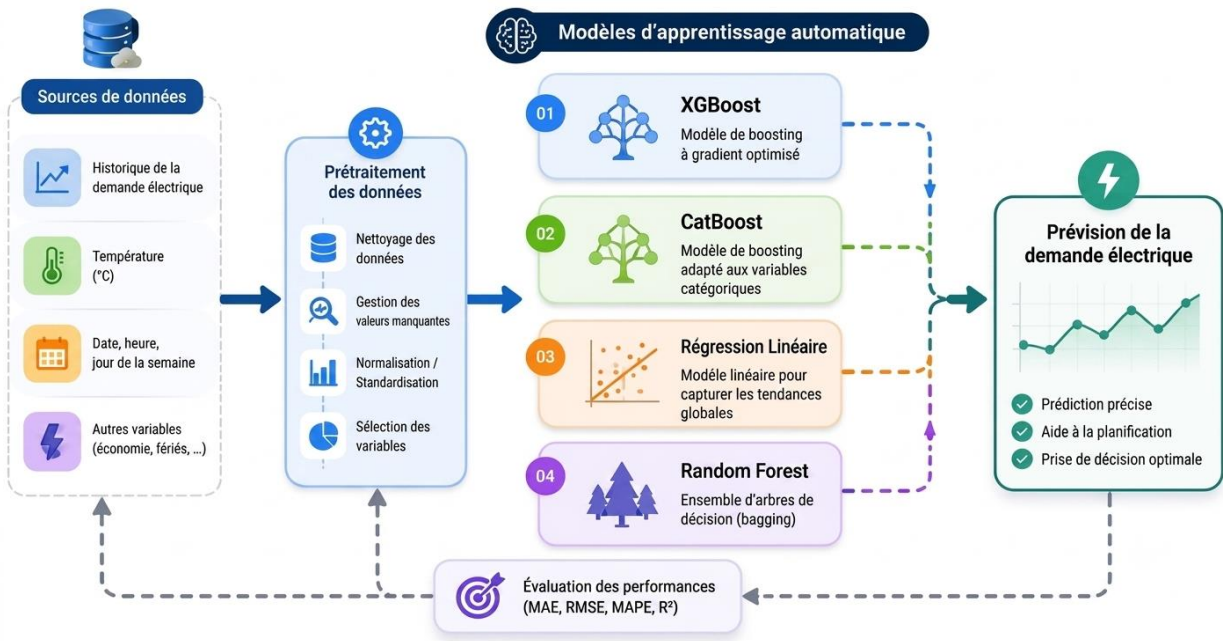


Figure 22: Organigramme du cadre de prédiction de notre étude

Le tableau présenté compare les performances de quatre modèles de régression (CatBoost, XGBoost, Random Forest et Régression Linéaire) en utilisant plusieurs métriques d'évaluation : MSE, RMSE, MAE et R^2 . L'analyse des résultats montre clairement que le modèle CatBoost est le plus performant parmi les modèles étudiés. En effet, il affiche les plus faibles valeurs d'erreur avec un MSE de 0.408103, un RMSE de 0.638829 et un MAE de 0.480083, ce qui indique une très bonne précision des prédictions. De plus, son coefficient de détermination R^2 atteint 0.986685, ce qui signifie qu'il explique environ 98.67 % de la variance des données, traduisant une excellente capacité d'ajustement. Le modèle XGBoost arrive en deuxième position avec des performances très proches de CatBoost. Il présente un MSE de 0.421658, un RMSE de 0.649352 et un MAE de 0.487119, ainsi qu'un R^2 de 0.986242. Bien que légèrement moins performant, il reste un modèle très fiable et constitue une alternative solide. Le modèle Random [82] Forest montre des performances intermédiaires. Ses erreurs sont plus élevées (MSE de 0.48818, RMSE de 0.698699, MAE de 0.523081) et son R^2 (0.984072) est légèrement inférieur à ceux de CatBoost et XGBoost. Cela indique une capacité de prédiction correcte, mais moins précise que les deux modèles précédents. En revanche, le modèle de Régression Linéaire présente les performances les plus faibles. Il enregistre les valeurs d'erreur les plus élevées avec un MSE de 0.620759, un RMSE de 0.787883 et un MAE de 0.62357, ainsi qu'un R^2 de 0.979746. Bien que ce R^2 reste relativement élevé, les erreurs importantes montrent que ce modèle s'adapte moins bien aux

données, probablement en raison de la complexité ou de la non-linéarité de celles-ci. En conclusion, CatBoost est le modèle à privilégier dans ce cas d'étude grâce à sa précision élevée et sa capacité explicative supérieure, suivi de près par XGBoost. Le modèle Random Forest offre des résultats acceptables mais moins performants, tandis que la Régression Linéaire apparaît comme le modèle le moins adapté pour ces données.

Tableau 2 : Évaluation comparative des performances de quatre modèles

Modèle	MSE	RMSE	MAE	R ²
CatBoost	0.408103	0.638829	0.480083	0.986685
XGBoost	0.421658	0.649352	0.487119	0.986242
Random Forest	0.48818	0.698699	0.523081	0.984072
Linear Regretion	0.620759	0.787883	0.62357	0.979746

La Figure 23 compare les coefficients de détermination (R^2) des différents modèles de prédiction. Tous les modèles présentent d'excellentes performances avec des valeurs de R^2 supérieures à 0,97. Le modèle CatBoost obtient le meilleur résultat ($R^2 = 0,986685$), suivi de très près par XGBoost ($R^2 = 0,986242$), ce qui confirme l'efficacité des algorithmes de boosting pour capturer les relations complexes entre les variables. Random Forest affiche également une très bonne précision ($R^2 = 0,984072$), tandis que la régression linéaire obtient la valeur la plus faible ($R^2 = 0,979746$), bien qu'elle conserve une capacité prédictive élevée. Ces résultats montrent que les modèles non linéaires, en particulier CatBoost et XGBoost, offrent les meilleures performances pour le problème étudié.

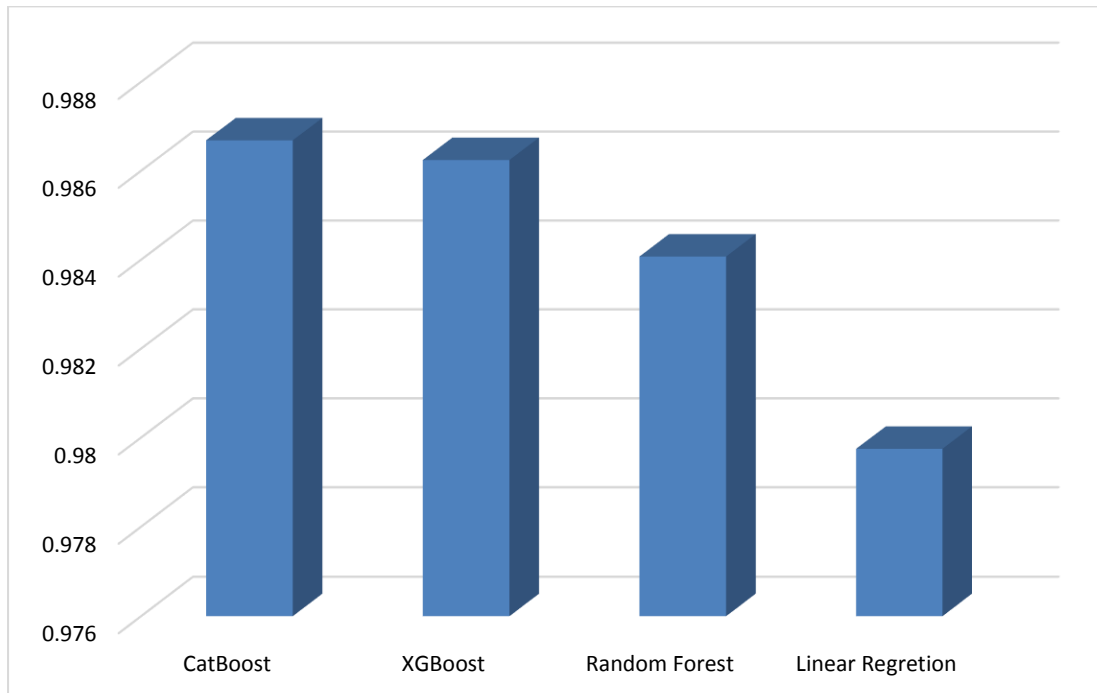


Figure 23 : Comparaison des coefficients de détermination des modèles utilisés

La Figure 24 présente une comparaison des performances des différents modèles à travers trois métriques d'erreur : MSE (Mean Squared Error), RMSE (Root Mean Squared Error) et MAE (Mean Absolute Error). Les résultats montrent que le modèle CatBoost obtient les erreurs les plus faibles sur l'ensemble des métriques, indiquant une meilleure précision de prédiction. XGBoost suit de très près avec des valeurs légèrement supérieures, tandis que Random Forest présente des erreurs intermédiaires. La régression linéaire affiche les performances les moins bonnes avec des erreurs nettement plus élevées. Cette figure met en évidence la supériorité des méthodes d'ensemble, en particulier les algorithmes de boosting, pour la modélisation du comportement étudié.

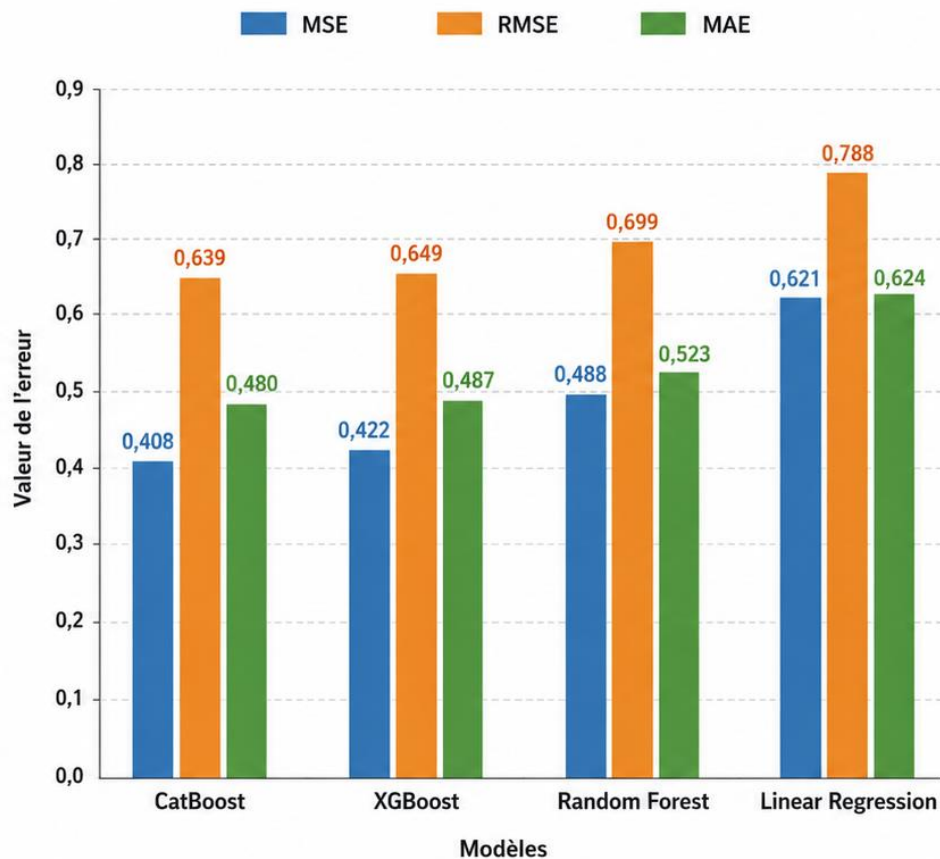


Figure 24 : Comparaison des erreurs (MSE, RMSE et MAE) sous forme d'histogramme groupé

La Figure 25 illustre le gain de performance des modèles d'apprentissage automatique par rapport à la régression linéaire en termes de réduction des erreurs (MSE, RMSE et MAE). Les résultats montrent que CatBoost offre le gain le plus important, suivi de XGBoost, puis de Random Forest. Plus précisément, les améliorations les plus significatives sont observées sur le MSE, ce qui confirme la capacité des modèles non linéaires à mieux capturer la complexité des données. Cette comparaison met en évidence l'intérêt des approches basées sur le boosting et les ensembles d'arbres, qui surpassent clairement le modèle linéaire de référence.

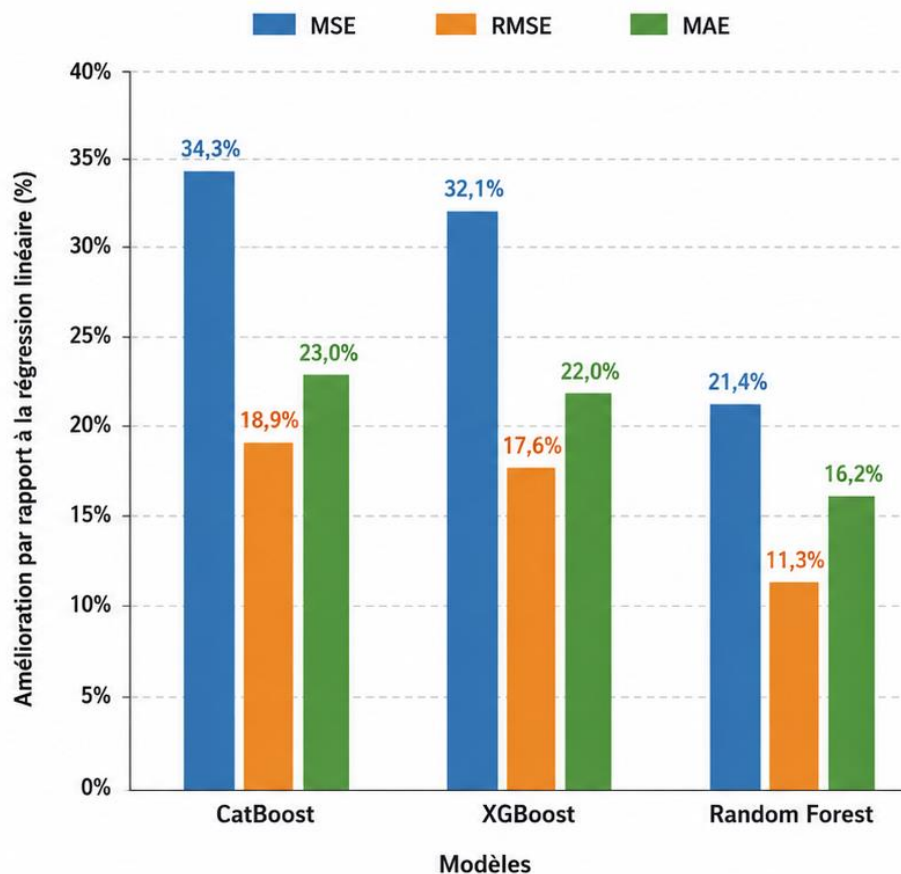


Figure 25 : Gain de performance des modèles de Machine Learning par rapport à la régression linéaire

4.4 Résultats du tracage du diagramme SHAP

Les valeurs SHAP (SHapley Additive Explanations) constituent une méthode d'interprétation permettant d'expliquer les prédictions réalisées par les modèles d'apprentissage automatique. Elles attribuent à chaque variable une contribution spécifique à la prédiction finale, ce qui permet de mesurer son importance et de comprendre comment elle influence les résultats du modèle. Les valeurs SHAP peuvent être appliquées à différents algorithmes tels que : la régression linéaire, Random forest, CatBoost et XGBoost.

La figure présentée correspond à un graphique SHAP Summary Plot du modèle XGBoost, illustrant l'impact des différentes variables explicatives sur la prédiction de la consommation électrique. Chaque ligne représente une variable utilisée dans le modèle, tandis que l'axe horizontal indique la contribution de cette variable à la prédiction. Les points colorés représentent les observations individuelles : les couleurs bleues correspondent aux faibles valeurs de la variable et les couleurs rouges aux valeurs élevées. La position des points sur

l'axe horizontal traduit l'effet de la variable sur la prédiction du modèle : vers la gauche pour une contribution négative et vers la droite pour une contribution positive. L'analyse du graphique montre que la variable `Peak_Hours_Usage_kWh` est la plus influente dans le modèle. Elle présente la plus grande dispersion des valeurs SHAP, s'étendant approximativement de -10 à +10. Les valeurs élevées de cette variable (points rouges) sont principalement situées du côté positif de l'axe, indiquant qu'une consommation importante pendant les heures de pointe augmente fortement la consommation électrique prédite. À l'inverse, les faibles valeurs (points bleus) se situent majoritairement du côté négatif, réduisant la prédiction du modèle. La variable `Household_Size` apparaît comme le deuxième facteur le plus important. Les valeurs élevées du nombre d'occupants d'un foyer tendent à produire des contributions positives, tandis que les faibles valeurs sont associées à des contributions négatives. Cela suggère qu'un ménage plus grand entraîne généralement une augmentation de la consommation énergétique. La variable `Has_AC`, qui indique la présence d'un système de climatisation, possède également une influence notable. Les observations associées à la présence de la climatisation contribuent davantage à l'augmentation des prédictions du modèle, tandis que son absence est liée à une diminution de la consommation estimée. Concernant `Avg_Temperature_C`, les valeurs élevées de température montrent une légère tendance à accroître la consommation électrique prédite. Cependant, son influence reste nettement inférieure à celle des variables précédentes. Les variables `Household_ID`, `DayOfWeek` et `Month` présentent des valeurs SHAP très proches de zéro. Leur impact sur les prédictions est donc relativement faible, indiquant qu'elles jouent un rôle secondaire dans le processus de décision du modèle.

Enfin, les variables `Year` et `Day` sont les moins importantes. Les points associés à ces variables sont fortement concentrés autour de zéro, ce qui traduit une contribution quasi nulle aux prédictions de consommation électrique. En conclusion, ce graphique SHAP met en évidence que le modèle XGBoost repose principalement sur la variable `Peak_Hours_Usage_kWh`, suivie de `Household_Size` et `Has_AC`, pour estimer la consommation électrique. Les autres variables ont une influence plus limitée. Cette représentation permet non seulement d'identifier les variables les plus importantes, mais également de comprendre la direction et l'intensité de leur contribution aux prédictions du modèle, offrant ainsi une meilleure interprétabilité des résultats obtenus.

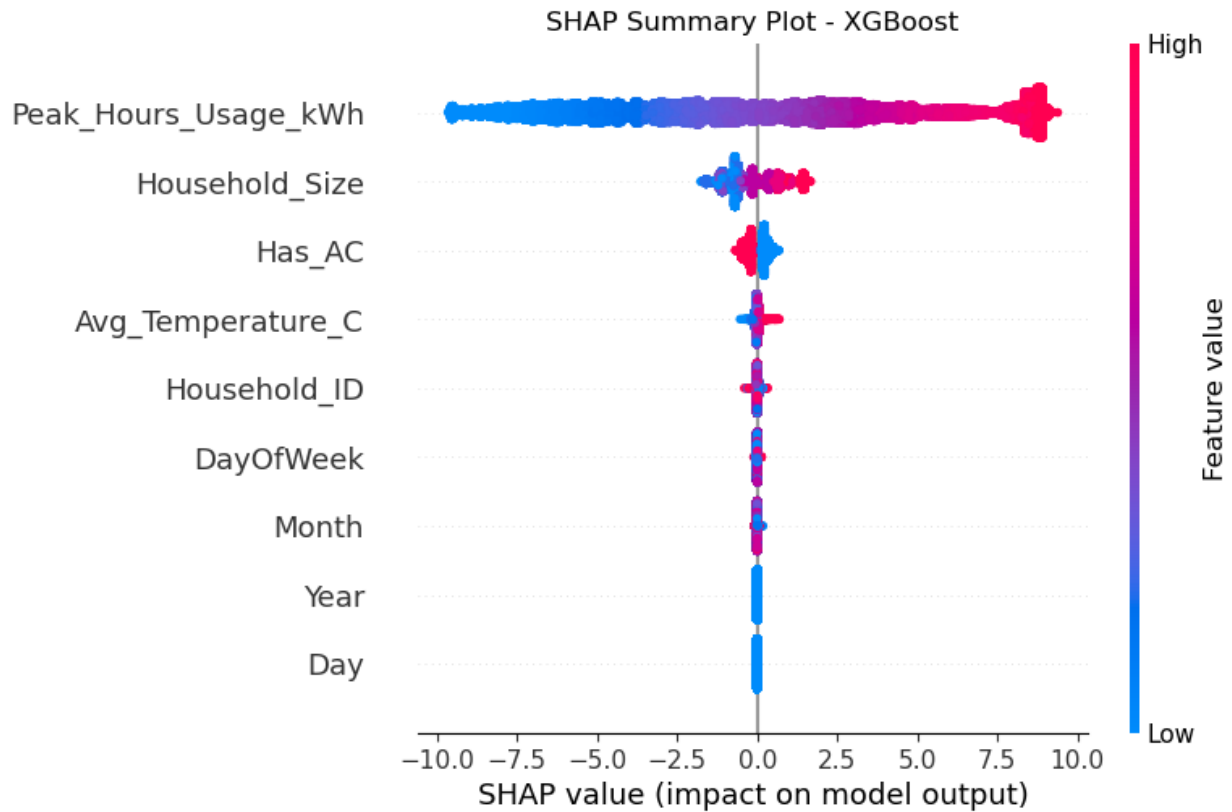


Figure 26 : Résultats du SHAP du modèle XGBoost

4.5 Conclusion

Dans ce chapitre, nous avons présenté et analysé les résultats obtenus à partir des différents modèles de Machine Learning utilisés pour la prévision de la consommation électrique. Les performances des modèles ont été comparées à l'aide des métriques MSE, RMSE, MAE et R^2 . Les résultats obtenus montrent que les modèles de boosting, notamment CatBoost et XGBoost, offrent les meilleures performances en termes de précision et de capacité de généralisation. L'analyse des valeurs SHAP a également permis d'identifier les variables les plus influentes dans le processus de prédiction, améliorant ainsi l'interprétabilité des modèles. Ces résultats confirment l'intérêt des techniques de Machine Learning dans le domaine de la prévision énergétique et mettent en évidence leur potentiel pour une meilleure gestion de la consommation électrique et le développement des systèmes énergétiques intelligents.

Conclusion Générale

La prévision de la consommation électrique constitue aujourd'hui un enjeu majeur dans le contexte de la transition énergétique et du développement des réseaux électriques intelligents. Une estimation précise de la demande énergétique permet d'assurer une meilleure planification des ressources, d'optimiser la production et la distribution de l'électricité, tout en contribuant à la réduction des coûts et à l'amélioration de l'efficacité énergétique. Dans ce cadre, l'intelligence artificielle et plus particulièrement les techniques de Machine Learning représentent des outils performants pour modéliser et prédire les comportements de consommation. L'objectif principal de ce travail était de développer et d'évaluer plusieurs modèles de Machine Learning afin de prédire la consommation électrique résidentielle à partir de différentes variables explicatives. Pour atteindre cet objectif, une étude approfondie a été menée en s'appuyant sur un jeu de données décrivant les caractéristiques des ménages ainsi que certains paramètres environnementaux influençant la consommation énergétique. Dans un premier temps, nous avons présenté les notions théoriques relatives à l'énergie électrique, aux séries temporelles ainsi qu'aux concepts fondamentaux du Machine Learning. Une étude bibliographique a ensuite permis d'examiner les travaux de recherche les plus récents dans le domaine de la prévision énergétique. Cette analyse a mis en évidence l'intérêt croissant des méthodes d'apprentissage automatique et leur supériorité par rapport aux approches statistiques traditionnelles pour la modélisation de phénomènes complexes et non linéaires. Par la suite, une méthodologie rigoureuse a été mise en place. Les données utilisées ont fait l'objet de plusieurs étapes de prétraitement comprenant le nettoyage des données, la transformation des variables temporelles, la normalisation ainsi que la séparation des données en ensembles d'apprentissage et de test. Ces différentes opérations ont permis d'améliorer la qualité des données et de garantir des conditions favorables au développement des modèles prédictifs. Quatre modèles de régression ont été implémentés et comparés, à savoir la Régression Linéaire, Random Forest, XGBoost et CatBoost. L'évaluation des performances a été réalisée à l'aide de plusieurs métriques, notamment l'erreur quadratique moyenne (MSE), la racine de l'erreur quadratique moyenne (RMSE), l'erreur absolue moyenne (MAE) et le coefficient de détermination (R^2). Les résultats obtenus ont montré que les modèles de boosting présentent les meilleures performances. Plus particulièrement, le modèle CatBoost s'est distingué par une excellente précision avec un MSE de 0,408103, un RMSE de 0,638829, un MAE de 0,480083 et un coefficient de détermination R^2 de 0,986685. Le modèle XGBoost a également fourni des résultats très satisfaisants, tandis que Random Forest a montré des performances intermédiaires. En revanche, la Régression Linéaire, bien qu'offrant une interprétation simple des résultats, s'est révélée moins adaptée à la

modélisation des relations complexes présentes dans les données. En complément de l'analyse des performances, l'interprétation du modèle XGBoost à l'aide des valeurs SHAP a permis d'identifier les variables les plus influentes dans le processus de prédiction. Les résultats ont montré que la consommation pendant les heures de pointe (Peak_Hours_Usage_kWh) constitue le facteur le plus déterminant, suivie de la taille du foyer (Household_Size) et de la présence d'un système de climatisation (Has_AC). Cette analyse a permis d'améliorer la compréhension du comportement des modèles et de renforcer leur interprétabilité. Les résultats obtenus démontrent ainsi l'efficacité des techniques de Machine Learning pour la prévision de la consommation électrique résidentielle. Ils confirment également que les méthodes de boosting, notamment CatBoost et XGBoost, représentent des solutions particulièrement performantes pour traiter des données complexes et fournir des prédictions précises. Ces modèles peuvent contribuer à une meilleure gestion énergétique, à la réduction des coûts de consommation et au développement des habitations intelligentes ainsi que des Smart Grids. Cependant, ce travail présente certaines limites. Les résultats obtenus dépendent fortement de la qualité et de la nature du jeu de données utilisé. De plus, seules quelques variables explicatives ont été prises en compte, alors que d'autres facteurs tels que l'humidité, la vitesse du vent, les habitudes de vie des occupants ou les tarifs de l'électricité peuvent également influencer la consommation énergétique. Comme perspectives futures, il serait intéressant d'utiliser des données réelles couvrant des périodes plus longues et d'intégrer un plus grand nombre de variables environnementales et comportementales. Il serait également pertinent d'explorer d'autres approches plus avancées basées sur le Deep Learning, telles que les réseaux de neurones récurrents (RNN), les modèles LSTM ou les architectures hybrides combinant plusieurs techniques d'intelligence artificielle. Ces améliorations pourraient permettre d'accroître davantage la précision des prévisions et de contribuer au développement de systèmes énergétiques intelligents, plus efficaces et plus durables.

Références

- [1] M. Waseem *et al.*, « Electrical Demand and its Flexibility in Different Energy Sectors », *Electr. Power Compon. Syst.*, vol. 48, n° 12-13, p. 1339-1361, août 2020, doi: 10.1080/15325008.2020.1854372.
- [2] Y. Demirel, « Energy and Energy Types », in *Energy: Production, Conversion, Storage, Conservation, and Coupling*, Y. Demirel, Éd., London: Springer, 2012, p. 27-70. doi: 10.1007/978-1-4471-2372-9_2.
- [3] J. L. Kirtley, *Electric Power Principles: Sources, Conversion, Distribution and Use*. John Wiley & Sons, 2020.
- [4] F. Harrou, A. Dorbane, et Y. Sun, « Enhancing photovoltaic system performance using advanced gradient boosting methods », in *2024 International Conference on Control, Automation and Diagnosis (ICCAD)*, mai 2024, p. 1-6. doi: 10.1109/ICCAD60883.2024.10553919.
- [5] F. Harrou, A. Dairi, A. Dorbane, B. Taghezouit, et Y. Sun, « Semi-supervised anomaly detection in photovoltaic systems under power tracking mode », *Energy Convers. Manag.*, vol. 352, p. 121114, mars 2026, doi: 10.1016/j.enconman.2026.121114.
- [6] « Power Generation | Springer Nature Link ». Consulté le: 9 mars 2026. [En ligne]. Disponible sur: https://link.springer.com/chapter/10.1007/978-3-030-47035-7_27
- [7] « The fossil fuel power plants technology », in *Materials for Ultra-Supercritical and Advanced Ultra-Supercritical Power Plants*, Woodhead Publishing, 2017, p. 1-49. doi: 10.1016/B978-0-08-100552-1.00001-4.
- [8] S. Sorrell, « Reducing energy demand: A review of issues, challenges and approaches », *Renew. Sustain. Energy Rev.*, vol. 47, p. 74-82, juill. 2015, doi: 10.1016/j.rser.2015.03.002.
- [9] « A review on hydro power plants and turbines | AIP Conference Proceedings | AIP Publishing ». Consulté le: 9 mars 2026. [En ligne]. Disponible sur: <https://pubs.aip.org/aip/acp/article-abstract/2418/1/030048/2822403/A-review-on-hydro-power-plants-and-turbines>
- [10] É. Larousse, « Centrales thermique et hydroélectrique – Média LAROUSSE ». Consulté le: 9 mars 2026. [En ligne]. Disponible sur: https://www.larousse.fr/encyclopedie/images/Centrales_thermique_et_hydro%C3%A9lectrique/1011749
- [11] « Fonctionnement d'une centrale thermique et d'une centrale hydroélectrique - Recherche Google ». Consulté le: 28 juin 2026. [En ligne]. Disponible sur: <https://www.google.com>
- [12] P. Bansal, E. Vineyard, et O. Abdelaziz, « Advances in household appliances- A review », *Appl. Therm. Eng.*, vol. 31, n° 17, p. 3748-3760, déc. 2011, doi: 10.1016/j.applthermaleng.2011.07.023.
- [13] M. Fahim et A. Sillitti, « Analyzing Load Profiles of Energy Consumption to Infer Household Characteristics Using Smart Meters », *Energies*, vol. 12, n° 5, p. 773, janv. 2019, doi: 10.3390/en12050773.
- [14] M. G. Fikru et L. Gautier, « The impact of weather variation on energy consumption in residential houses », *Appl. Energy*, vol. 144, p. 19-30, avr. 2015, doi: 10.1016/j.apenergy.2015.01.040.
- [15] « Evolution de la facture d'électricité en fonction des températures en ile-de-France (Mai 2020 – Avril 2021, comparée a Mai 2019 – Avril 2021) - Recherche Google ». Consulté le: 28 juin 2026. [En ligne]. Disponible sur: <https://www.google.com/>

- [16] L. Enshen, « Research on the influence of air humidity on the annual heating or cooling energy consumption », *Build. Environ.*, vol. 40, n° 4, p. 571-578, avr. 2005, doi: 10.1016/j.buildenv.2004.06.017.
- [17] X. Xu, B. Xiao, et C. Z. Li, « Critical factors of electricity consumption in residential buildings: An analysis from the point of occupant characteristics view », *J. Clean. Prod.*, vol. 256, p. 120423, mai 2020, doi: 10.1016/j.jclepro.2020.120423.
- [18] K. Gaur, H. Kumar, R. P. K. Agarwal, K. V. S. Baba, et S. K. Soonee, « Analysing the electricity demand pattern », in *2016 National Power Systems Conference (NPSC)*, déc. 2016, p. 1-6. doi: 10.1109/NPSC.2016.7858969.
- [19] I. Ihsane, « Préviation à court terme et gestion des consommations d'énergie électrique dans l'habitat », Theses, nantes université, 2020. Consulté le: 9 mars 2026. [En ligne]. Disponible sur: <https://hal.science/tel-03859920>
- [20] F. Melzi, « Fouille de données pour l'extraction de profils d'usage et la prévision dans le domaine de l'énergie », Theses, Université Paris-Est, 2018. Consulté le: 9 mars 2026. [En ligne]. Disponible sur: <https://theses.hal.science/tel-02127788>
- [21] « Baromètre de la consommation d'énergie en France ». Consulté le: 9 mars 2026. [En ligne]. Disponible sur: <https://www.hellowatt.fr/suivi-consommation-energie/consommation-electrique/barometre-consommation-france>
- [22] H. Watt, « Baromètre de la consommation d'énergie en France », Hello Watt. Consulté le: 12 mai 2026. [En ligne]. Disponible sur: <https://www.hellowatt.fr/suivi-consommation-energie/consommation-electrique/barometre-consommation-france>
- [23] E. M. P. Hounton, « Prédiction pour l'optimisation de la consommation électrique résidentielle à l'aide de l'apprentissage automatique sur des données électriques et météorologiques », masters, Université du Québec à Chicoutimi, Chicoutimi, 2025. Consulté le: 9 mars 2026. [En ligne]. Disponible sur: <https://constellation.uqac.ca/id/eprint/10304/>
- [24] A. Bellahsen, *L'intelligence artificielle au service de l'optimisation de l'énergie électrique dans un réseau intelligent*. Ecole Polytechnique, Montreal (Canada), 2020.
- [25] G. Cerri, M. Gazzino, F. A. Iacobone, et A. Giovannelli, « Optimum Planning of Electricity Production », *J. Eng. Gas Turbines Power*, vol. 131, n° 061801, juill. 2009, doi: 10.1115/1.3098429.
- [26] S. Djeddi, « Efficacité énergétique levier de la transition énergétique: le rôle de l'efficacité énergétique dans la transformation du système énergétique », *Le Manager*, vol. 9, n° 3, p. 184-207, 2022.
- [27] D. Six, « Analyse statistique des distributions des séries de bilans de masse des glaciers alpins et des calottes polaires de l'hémisphère nord », Theses, Université Joseph-Fourier - Grenoble I, 2000. Consulté le: 7 avril 2026. [En ligne]. Disponible sur: <https://theses.hal.science/tel-00701276>
- [28] I. Tall, « Introduction aux Séries temporelles avec R ».
- [29] N. Chapados, « Critères d'optimisation d'algorithmes d'apprentissage en gestion de portefeuille », 2000, Consulté le: 7 avril 2026. [En ligne]. Disponible sur: <http://hdl.handle.net/1866/30770>
- [30] N. A. Gershenfeld et A. S. Weigend, « The Future of Time Series: Learning and Understanding », in *Pattern Formation In The Physical And Biological Sciences*, CRC Press, 1997.
- [31] Y. Hmamouche, « Prédiction des Séries Temporelles Larges ».
- [32] G. Belbachir, « Analyse des Séries Temporelles », déc. 2024, Consulté le: 28 avril 2026. [En ligne]. Disponible sur: <https://dSPACE.ummto.dz/handle/ummto/25862>

- [33] Louis, « Les séries temporelles avec Python (4/4) - Analyse et modélisation », Le blog. Consulté le: 24 avril 2026. [En ligne]. Disponible sur: <https://blog.statoscop.fr/timeseries-4.html>
- [34] T. Abergel et M. Jordan, « Chauffage et climatisation : enjeux et opportunités en France, en Europe et dans le reste du monde », *Ann. Mines - Responsab. Environ.*, vol. 95, n° 3, p. 46-49, juill. 2019, doi: 10.3917/re1.095.0046.
- [35] E. Ramasso, « Méthodologies d'analyse de séries temporelles sous incertitudes aléatoires et épistémiques pour le suivi et le pronostic de l'état de systèmes et structures - De l'estimation d'une cinétique d'endommagement à son contrôle ».
- [36] « blog.statoscop.fr/tag/series-temporelles.html ». Consulté le: 24 avril 2026. [En ligne]. Disponible sur: <https://blog.statoscop.fr/tag/series-temporelles.html>
- [37] A. Quartier-La-Tente, « Estimation en temps réel de la tendance-cycle à l'aide de moyennes mobiles : apports dans l'analyse conjoncturelle », phdthesis, Nantes Université, 2025. Consulté le: 28 avril 2026. [En ligne]. Disponible sur: <https://theses.hal.science/tel-05465892>
- [38] A. Evrard et S. Persico, « Chapitre 10. Entre relance verte et variable d'ajustement : les effets contradictoires de la Grande Récession sur les politiques environnementales », in *Les Politiques publiques dans la crise*, Presses de Sciences Po, 2021, p. 297-329. doi: 10.3917/scpo.hasse.2021.01.0297.
- [39] « Informed Machine Learning – A Taxonomy and Survey of Integrating Prior Knowledge into Learning Systems | IEEE Journals & Magazine | IEEE Xplore ». Consulté le: 28 avril 2026. [En ligne]. Disponible sur: <https://ieeexplore.ieee.org/abstract/document/9429985>
- [40] X. Deloffre, « Machine learning → Définition & fonctionnement, histoire », Facem Web. Consulté le: 29 avril 2026. [En ligne]. Disponible sur: <https://facemweb.com/blog/notoriete/machine-learning-definition/>
- [41] Z. Yousnadj, « Optimisation de l'énergie électrique dans un réseau intelligent en utilisant l'intelligence artificielle », 2025. doi: 10.13140/RG.2.2.23301.61929.
- [42] R. Tighilt, « Étude comparative d'algorithmes d'apprentissage profond pour la prédiction de futures courses d'épicerie ».
- [43] J.-S. Chou et D.-S. Tran, « Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders », *Energy*, vol. 165, p. 709-726, déc. 2018, doi: 10.1016/j.energy.2018.09.144.
- [44] C. Deb, F. Zhang, J. Yang, S. E. Lee, et K. W. Shah, « A review on time series forecasting techniques for building energy consumption », *Renew. Sustain. Energy Rev.*, vol. 74, p. 902-924, juill. 2017, doi: 10.1016/j.rser.2017.02.085.
- [45] M. FLORENT, « Neurones artificiels et neurones biologiques », Tree Learning. Consulté le: 26 mai 2026. [En ligne]. Disponible sur: <https://www.tree-learning.fr/plateforme-lms-elearning/neurones-artificiels/>
- [46] K. Amasyali et N. M. El-Gohary, « A review of data-driven building energy consumption prediction studies », *Renew. Sustain. Energy Rev.*, vol. 81, p. 1192-1205, janv. 2018, doi: 10.1016/j.rser.2017.04.095.
- [47] M. Bourdeau, X. qiang Zhai, E. Nefzaoui, X. Guo, et P. Chatellier, « Modeling and forecasting building energy consumption: A review of data-driven techniques », *Sustain. Cities Soc.*, vol. 48, p. 101533, juill. 2019, doi: 10.1016/j.scs.2019.101533.
- [48] « (PDF) Optimisation des Ressources Énergétiques grâce à la Prédiction de la Consommation Électrique : Approche Analytique et Prédicative basée sur l'Intelligence Artificielle », ResearchGate. Consulté le: 26 mai 2026. [En ligne]. Disponible sur: https://www.researchgate.net/publication/385592027_Optimisation_des_Ressources_En

- ergetiques_grace_a_la_Prediction_de_la_Consommation_Electrique_Approche_Analytique_et_Predictive_basee_sur_l'Intelligence_Artificielle
- [49] « Electricity load forecasting: a systematic review | Journal of Electrical Systems and Information Technology | Springer Nature Link ». Consulté le: 12 mai 2026. [En ligne]. Disponible sur: https://link.springer.com/article/10.1186/s43067-020-00021-8?utm_source=chatgpt.com
- [50] C. Kuster, Y. Rezgui, et M. Mourshed, « Electrical load forecasting models: A critical systematic review », *Sustain. Cities Soc.*, vol. 35, p. 257-270, nov. 2017, doi: 10.1016/j.scs.2017.08.009.
- [51] Y. Sun, F. Haghighat, et B. C. M. Fung, « A review of the-state-of-the-art in data-driven approaches for building energy prediction », *Energy Build.*, vol. 221, p. 110022, août 2020, doi: 10.1016/j.enbuild.2020.110022.
- [52] K. P. Amber, R. Ahmad, M. W. Aslam, A. Kousar, M. Usman, et M. S. Khan, « Intelligent techniques for forecasting electricity consumption of buildings », *Energy*, vol. 157, p. 886-893, août 2018, doi: 10.1016/j.energy.2018.05.155.
- [53] A. Gellert, U. Fiore, A. Florea, R. Chis, et F. Palmieri, « Forecasting Electricity Consumption and Production in Smart Homes through Statistical Methods », *Sustain. Cities Soc.*, vol. 76, p. 103426, janv. 2022, doi: 10.1016/j.scs.2021.103426.
- [54] M. Bilal, H. Kim, M. Fayaz, et P. Pawar, « Comparative Analysis of Time Series Forecasting Approaches for Household Electricity Consumption Prediction », 3 juillet 2022, *arXiv*: arXiv:2207.01019. doi: 10.48550/arXiv.2207.01019.
- [55] M. Peplinski, B. Dilkina, M. Chen, S. J. Silva, G. A. Ban-Weiss, et K. T. Sanders, « A machine learning framework to estimate residential electricity demand based on smart meter electricity, climate, building characteristics, and socioeconomic datasets », *Appl. Energy*, vol. 357, p. 122413, mars 2024, doi: 10.1016/j.apenergy.2023.122413.
- [56] G. S. Ramnath, R. Harikrishnan, S. M. Muyeen, et K. Kotecha, « Household electricity consumption prediction using database combinations, ensemble and hybrid modeling techniques », *Sci. Rep.*, vol. 14, n° 1, p. 22891, oct. 2024, doi: 10.1038/s41598-024-57550-9.
- [57] A. Jain et S. C. Gupta, « Evaluation of electrical load demand forecasting using various machine learning algorithms », *Front. Energy Res.*, vol. 12, juin 2024, doi: 10.3389/fenrg.2024.1408119.
- [58] M. Baek et Y. Seo, « Hybrid forecasting of university electricity demand using time series and deep learning », *Energy Build.*, vol. 347, p. 116400, nov. 2025, doi: 10.1016/j.enbuild.2025.116400.
- [59] A. Muqtadir, B. Li, Z. Ying, C. Songsong, et S. N. Kazmi, « Nowcasting the next hour of residential load using boosting ensemble machines », *Sci. Rep.*, vol. 15, n° 1, p. 7157, févr. 2025, doi: 10.1038/s41598-025-91767-6.
- [60] J.-S. Chou et D.-S. Tran, « Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders », *Energy*, vol. 165, p. 709-726, déc. 2018, doi: 10.1016/j.energy.2018.09.144.
- [61] A. Hossain *et al.*, « Energy Demand Forecasting Using Machine Learning: Optimizing Smart Grid Efficiency with Time-Series Analytics », *J. Environ. Agric. Stud.*, vol. 5, n° 1, p. 26-42, févr. 2024, doi: 10.32996/jeas.2024.5.1.4.
- [62] M. Baek et Y. Seo, « Hybrid forecasting of university electricity demand using time series and deep learning », *Energy Build.*, vol. 347, p. 116400, nov. 2025, doi: 10.1016/j.enbuild.2025.116400.
- [63] « Hybrid forecasting of university electricity demand using time series and deep learning », *Energy Build.*, vol. 347, p. 116400, nov. 2025, doi: 10.1016/j.enbuild.2025.116400.

- [64] A. Gellert, U. Fiore, A. Florea, R. Chis, et F. Palmieri, « Forecasting Electricity Consumption and Production in Smart Homes through Statistical Methods », *Sustain. Cities Soc.*, vol. 76, p. 103426, janv. 2022, doi: 10.1016/j.scs.2021.103426.
- [65] Y. Wu, « Linear regression in machine learning », in *International Conference on Statistics, Applied Mathematics, and Computing Science (CSAMCS 2021)*, SPIE, avr. 2022, p. 1253-1264. doi: 10.1117/12.2628053.
- [66] J.-S. Chou et D.-S. Tran, « Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders », *Energy*, vol. 165, p. 709-726, déc. 2018, doi: 10.1016/j.energy.2018.09.144.
- [67] A. Hossain *et al.*, « Energy Demand Forecasting Using Machine Learning: Optimizing Smart Grid Efficiency with Time-Series Analytics », *J. Environ. Agric. Stud.*, vol. 5, n° 1, p. 26-42, févr. 2024, doi: 10.32996/jeas.2024.5.1.4.
- [68] J. Bedi et D. Toshniwal, « Deep learning framework to forecast electricity demand », *Appl. Energy*, vol. 238, p. 1312-1326, mars 2019, doi: 10.1016/j.apenergy.2019.01.113.
- [69] « Household Energy Consumption ». Consulté le: 29 avril 2026. [En ligne]. Disponible sur: <https://www.kaggle.com/datasets/samxsam/household-energy-consumption>
- [70] « Household Energy Consumption ». Consulté le: 29 juin 2026. [En ligne]. Disponible sur: <https://www.kaggle.com/datasets/samxsam/household-energy-consumption>
- [71] M. M. Koné, « Traitement des données de consommation d'électricité acquis d'un dispositif autonome de mesure préalablement conçu en vue de mieux gérer la demande », masters, École de technologie supérieure, 2026. Consulté le: 29 avril 2026. [En ligne]. Disponible sur: <https://espace.etsmtl.ca/id/eprint/3865/>
- [72] « Régression linéaire - expliquée simplement ». Consulté le: 10 juin 2026. [En ligne]. Disponible sur: <https://numiqo.fr/tutorial/linear-regression>
- [73] Z. M'Hamedi, C. Dartigues-Pallez, et T. Bell, « RFIViz : Random Forest Interactive Visualisation, un outil simple pour comprendre les modèles », présenté à EXPLAIN'AI 2024, janv. 2024. Consulté le: 10 juin 2026. [En ligne]. Disponible sur: <https://hal.science/hal-04906235>
- [74] « Fig. 3. Random Forest algorithms' structure [11] », ResearchGate. Consulté le: 10 juin 2026. [En ligne]. Disponible sur: https://www.researchgate.net/figure/Random-Forest-algorithms-structure-11_fig1_375676525
- [75] insightsoftware, « Les meilleurs modèles et algorithmes d'analyse prédictive à connaître », insightsoftware. Consulté le: 29 avril 2026. [En ligne]. Disponible sur: <https://insightsoftware.com/fr/blog/top-5-predictive-analytics-models-and-algorithms/>
- [76] D. Team, « XGBoost Tutorial - What is XGBoost in Machine Learning? », DataFlair. Consulté le: 10 juin 2026. [En ligne]. Disponible sur: <https://data-flair.training/blogs/xgboost-tutorial/>
- [77] « CatBoost in Machine Learning - GeeksforGeeks ». Consulté le: 10 juin 2026. [En ligne]. Disponible sur: <https://www.geeksforgeeks.org/machine-learning/catboost-ml/>
- [78] « CatBoost in Machine Learning: A Detailed Guide ». Consulté le: 10 juin 2026. [En ligne]. Disponible sur: <https://www.igmguru.com/blog/catboost>
- [79] « Prétraitement et analyse de données ».
- [80] « Nettoyage et traitement des données | PDF | Données | Informatique ». Consulté le: 10 juin 2026. [En ligne]. Disponible sur: <https://fr.scribd.com/document/736569996/Nettoyage-de-Donnees>
- [81] « Données numériques : normalisation | Machine Learning », Google for Developers. Consulté le: 10 juin 2026. [En ligne]. Disponible sur: <https://developers.google.com/machine-learning/crash-course/numerical-data/normalization?hl=fr>

[82] « Mémento sur l'énergie - édition 2015 ». Consulté le: 17 décembre 2025. [En ligne]. Disponible sur: <https://www.cea.fr/english/Documents/scientific-and-economic-publications/memento-2015.pdf>